MICROWAVES and RADAR ELECTRONICS

By Ernest C. Pollard

and

Julian M. Sturtevant

MICROWAVES AND RADAR ELECTRONICS

By Ernest C. Pollard

and

William L. Davidson, Jr.

APPLIED NUCLEAR PHYSICS

MICROWAVES and RADAR ELECTRONICS

ERNEST C. POLLARD
Professor of Physics

JULIAN M. STURTEVANT
Associate Professor of Chemistry

both at Yale University

1948 • JOHN WILEY & SONS, INC., NEW YORK
CHAPMAN & HALL, LTD., LONDON

Second Printing, May, 1951

Printed in the United States of America

PREFACE

As a result of the great amount of publicity given to the wartime development of radar it is now widely realized that this development involved striking advances in several fields of physics and technology. The current scientific literature of physics and chemistry, and the technological literature of communications and other engineering fields, show clearly that these advances have numerous interesting and important applications in problems far removed from those of radar. A visitor to a physical research laboratory today cannot avoid being impressed by the profound change in the kind of equipment which is being built and used. Large cathode ray tube indicators, microwave cavities and waveguides, electronic computers, and pulsed circuits of many types are now commonplace, whereas eight years ago such devices were rarely used.

In order that these new techniques and manners of thought may be exploited to the fullest extent in the various fields of science and engineering there is need for an overall presentation in compact form which will enable one who had no opportunity to participate in their development to appreciate their potentialities. This book has such a presentation for its purpose.

It is obvious that a full coverage of this material is impossible in a single volume of moderate length. It is also true that a detailed and thorough treatment would require the collective effort of numerous authors. We have attempted here to present a survey broad enough to give a general view of the field and a background on which the reader can base further intensive study of any phases of particular interest to him. The presentation is partly explanatory and partly descriptive. A short introduction outlines the special knowledge of electromagnetic fields which is necessary to an understanding of microwaves. Apart from this, the material presented needs no special introductory knowledge other than elementary physics, a modest amount of mathematics, and a comprehension of the elementary facts concerning electron tubes. The

book should thus be satisfactory for use in an advanced undergraduate, or graduate, course in physics or electrical engineering dealing with the subjects covered.

Special attention has been given to the applications of the subject matter of this book to the scientific fields within which the authors have some competency, namely, physics and chemistry. Progress in these applications is at present very rapid, so that by the time this book appears in print there will no doubt be many interesting new applications in the periodical literature. However. it is hoped that the illustrations given here will show the diversity and importance of these applications.

Most of the topics considered in this book are given a detailed and authoritative treatment in the twenty-seven-volume *Radiation Laboratory Series*. Specific references to this series have not usually been included in the text because only a few of the volumes had appeared at the time the manuscript was completed.

Our experience in radar was gained at the Radiation Laboratory of the Massachusetts Institute of Technology. Very important contributions to the radar program were made at numerous other laboratories in this country and abroad, and it is probable that these contributions have to some extent been slighted in this book. This is an unavoidable result of the wartime conditions under which the radar program was carried out. Another unfortunate result of these conditions is that it is quite out of the question to give proper credit even to those who made outstanding individual contributions to the development of radar. We wish, however, to acknowledge our great debt to our many colleagues at the Radiation Laboratory for the stimulus in research and discussion which rendered the period of service within its walls so rich and permanent an experience.

E. C. P.
J. M. S.

New Haven, Connecticut
August, 1948

CONTENTS

CHAPTER 1

ELECTROMAGNETIC FIELDS AND MICROWAVES

To gain some measure of control over Nature we are required to deal with extremes. The velocity of light is 30 billion centimeters per second, the frequency of a light wave is 10^{10} million cycles per second. These illustrate the stock in trade with which Nature operates. If we are to have some influence in our own material affairs, we have to learn how to manage such extreme quantities. Progress in this art is a combination of the mastery of new methods and an assimilation of the experimental results by theory. The two do not always go together, yet they are essential to one another. This book deals largely with technique. The technique stems from microwave radar. Radar requires high power electromagnetic radiation, short pulses, fast sweeps, and special indicators. Microwave radar, in addition, requires the radiation to be of wavelengths in the centimeter region. It is thus natural that the development of microwave radar carried with it a great advance in technique on a broad front. This advance is now showing gains in many research fields other than radar.

Electrons and Fields

Almost the whole of the subject matter to be described depends on the fact that electrons are very light and respond to electric fields without much inertia effect. It is worth a moment to consider what happens to an electron which finds itself in an electric field of 1 volt per centimeter. The acceleration is roughly 2×10^{12}

times that of gravity. In the first microsecond the electron travels 10 meters. There is therefore no question about the ability of electrons to follow very fast processes. This has been known for a long time but has had, until 1940, surprisingly little exploitation. Electrons had been used in vacuum tubes as inertialess current carriers which can be controlled by small voltages, but no special speeds were developed. At the time when cyclotrons were being first constructed, the 10-megacycle frequency used with them was felt to be quite high. The fastest timing operations before the war of which the authors are aware, the work of Dunnington and White on spark onset, were done with Kerr cells and used light as the indicating agent.

This exploitation of fast time operation has begun with the work on radar. It is exploited in the klystron and the magnetron to give very high frequency electromagnetic oscillations, going as high as 100,000 megacycles. The fastest cathode ray tube sweeps cover the face of the tube in less than a microsecond and are already being used to study gaseous discharges in experiments which are in the same field as the work of Dunnington and White. Other less visible fast operations are going on. For example, an automatic range tracking device samples voltage due to an echo response in a section of time located within a microsecond. The sampling is at the rate of once in half a millisecond, and the result of the sampling can be made to hold a marker accurately at the position of the echo from a shell in flight. It is possible, by adapting some of the rapidly switched tubes in a device of this kind, to complete a solution of a simple equation in a very small fraction of a second.

This kind of technical advance is not made by wish alone. A pulsed magnetron must be pulsed by the right voltage, and the right current must be supplied. A switching circuit will not operate at a microsecond rate unless the circuit capacitance is below a certain figure and the tube constants are right. One may wire up a 30-megacycle amplifier and amplify nothing at all. All these features of technique are perfectly reasonable once the whole process is carefully understood, whereupon the secondary effects operating are taken into account. The engineering of radar has brought under control the following specific fields: high power cavity magnetrons from 1 centimeter wavelength to 30 centimeters; klystrons, velocity modulation tubes which are able to

deliver continuous power from a milliwatt up to a number of watts in the wavelength range from ½ centimeter to 30 centimeters; triode oscillators which will deliver up to a kilowatt at 30 centimeters and several watts at 6 centimeters. The technique of delivering microwave power from the source to some objective, like an antenna, has been reduced to a relatively simple process. Means for switching channels on or off by high frequency gas discharges have been devised.

In conjunction with this have been the excellent engineering of crystal detectors which approach theoretical sensitivity, intermediate frequency amplifiers at 30 to 60 megacycles, video amplifiers of 10-megacycle bandpass, sweep circuits for oscilloscopes, computers to transfer angular information from one place to another, switching circuits to turn amplifiers on or off or intensify oscilloscopes. Means which have great ultimate sensitivity have been devised for the study of variable modulation.

This brief listing indicates the technical scope of the microwave radar development. In this chapter the consideration of the general properties of electromagnetic fields is followed by an account of microwaves.

1·1 ELECTRIC AND MAGNETIC FIELDS

Microwaves, of wavelength between 1 millimeter and, say, 30 centimeters, occupy a position between the definitely man-made oscillations of long wave radio and the atomic radiations of light and heat. Very short microwaves can be used to study very low frequency molecular transitions. Because of this position, overlapping two sciences, the language of microwaves is somewhat unusual. It is a composite of physical language in terms of fields and engineering language in terms of impedance. Both are of great value. However, in microwaves, a larger share of reasoning is done in terms of fields and therefore their properties need consideration.

The more familiar electromagnetic waves, such as are used for broadcast purposes, are produced by visible circuit elements. Thus one can see the coil and condenser in the tank circuit of an oscillator which can produce radiation at 10 megacycles. This ability to see coiled wires and condenser plates disappears at microwave frequencies, and the whole art of microwaves is de-

pendent on an understanding of the basic relations of electricity and magnetism, of which circuit relations are only a particular case. The method of thought which has grown up around circuits is so valuable and powerful that it has been retained in much microwave development. Nevertheless, to understand wave propagation, radiation, "skin effect," and waveguides it is essential to be able to call on the fundamental equations of the electromagnetic field. Therefore in this chapter we review these equations, as simply as we can; and because present and future microwave experiments are related to fundamental electrical quantities we carry the treatment all the way back to the properties of charges and fields.

To do this, a choice of units is necessary. The units chosen here are Gaussian units; that is, electric fields are in statvolts per centimeter, magnetic fields in oersteds or gauss, electric charge in statcoulombs, and current in statamperes. These units are chosen because they appear to us to be the simplest for the description of fundamental processes, not because they are necessarily the best for engineering problems, or because the equations take on a specially compact form. In Appendix 3 we discuss units further and show how the relations given in this chapter appear in the mks system.

Electric and magnetic phenomena comprise electric and magnetic fields which are strongly interrelated with one another. To describe these accurately one should present the complete picture at once, for the whole picture is needed to describe radiation. This is, nevertheless, the more difficult way. It is more usual to describe the purely electrostatic field which is produced by a stationary electric charge and to develop the complete picture from that starting point. This we can now do.

Field Near an Electric Charge—Coulomb's Law

It is known that the vast majority of natural processes and material structures can be explained in terms of elementary particles, notably electrons and protons. The reason for the existence of these and their precise nature are still not clear, but most of the effects they produce can be accurately described. Much the biggest effect is the presence of an electric *field* around each charged particle. This field is determined by a total of four quantities, of

which we first consider three: the charge on the particle, the magnitude of the distance from the particle, and the direction from the point considered to the particle. Another factor is also of great importance: the time which the particle has spent at the location under consideration. We assume at first that this time has been very long; later we consider the effects when the time has been short.

We now describe the field. It is entirely unnoticed until a second charged particle is brought to the point considered. The field then becomes observable in terms of a mechanical force which is used as a measure of the field. If we agree to call the field[1] $\mathbf{E}$ and if the magnitude of the charge on the second particle is e, we use the relation

$$\text{Force } \mathbf{F} = e\mathbf{E}$$

to measure the electric field $\mathbf{E}$. Notice that the force has direction (that is, is a vector quantity), and so does $\mathbf{E}$. The direction is along the line to the original particle.

It was found by Coulomb and verified by Cavendish and Maxwell that $\mathbf{E}$ varies as $1/r^2$, where r is the distance between the particles. This can be made the basis for a definition of a unit charge as the charge which repels a like charge with a force of 1 dyne at 1 centimeter. With this form of definition we have for the *magnitude* of the field at a distance r from a charge e

$$E = \frac{e}{r^2} \tag{1·1}$$

and the action of the field $\mathbf{E}$ is along the line from the charge to the point considered. This is a statement of *Coulomb's law.*

Electrons and protons are nearly always present by the million at least. In addition the motion of charges is complicated by their having to move through matter containing enormous numbers of atoms. We defer discussing the effect of this complication until later. At present we call attention to the fact that protons and electrons have charges equal in magnitude but opposite in sign. Reversal of the sign of course reverses the direction of the force acting. It also makes possible the *neutrality* of matter. The

[1] A bold-faced character indicates a vector, meaning a quantity having magnitude and direction.

amounts of plus and minus charge are equal to a very close approximation. This equality is perhaps the most exact property of matter.

Energy in the Electric Field

The existence of a force between electric charges challenges us to try to devise an interpretation. The simplest solution is to class it with gravitation as an effect which is itself basic. This view was held until Faraday came to the conclusion that both electric effects and magnetic effects can better be understood if they are explained in terms of a medium. Faraday used a rather specific mechanical analogue as an aid in interpretation, and this was developed further by Maxwell. Today the idea of an ether capable of physical distortion is very hard to maintain, but the idea introduced by Faraday that the electric effects are a property of the space around the charge rather than solely in the charge itself is firmly held. One thinks simply of an electric field around the charge and not of the field as a stress in an elastic medium. *The important idea is the reality of the field.*

If we take this view it is in order to find some expression for the energy per unit volume in a space containing a field $\mathbf{E}$. With our basic assumption, that the field is real and carries energy, it presumably will not matter how we find this energy. We can imagine a field created in some simple way, for example by confining it to a definite volume between two plates, and observe the energy necessary to put it there. Consider the case of two plates, distance d centimeters apart with a surface charge density S, plus on one plate, minus on the other. The electric field in such a case is $E = 4\pi S$, and the work done in putting the surface charge in place is the work done in taking the plus charge from one plate to the other, so leaving an equal minus charge behind. This work is the product of the average of a field which starts at zero and ends up as E, namely $E/2$, times the charge times the distance. For each square centimeter of surface the charge is S statcoulombs; therefore the work done is $2\pi S^2 d$. Since it is more fundamental to express this in terms of field strength we substitute $S = E/4\pi$ which gives for the energy in the space bounded by 1 square centimeter and the distance d the value $E^2 d/8\pi$. The energy in ergs per cubic centimeter is then $E^2/8\pi$.

Effect of a Changing Electric Field

There is no *a priori* reason why the only observable effect of an electric field should be the production of a force on a stationary charge. Indeed, if we are to explain light and radio waves as being electrical in nature, it is quite clear that this cannot be so because wave motion is not static in character. We therefore expect that additional effects will be observed when the electric field is changing. The correct description of these effects can hardly be expected to be quite as simple as the description of the electric field itself, since we have to be concerned with the size of the field, its direction, and its rate of change. Moreover, it may well be that the varying electric field is without effect on a stationary charge but affects a moving charge. Hence a certain amount of intricacy must be expected.

The existence of some added effect in moving charges (and hence changing fields) is seen easily by arranging for charges to move parallel to one another in the same direction. Like charges should repel, yet the fact is that the moving charges repel each other less while moving. This action of moving charges on one another is called *magnetism*. It was studied empirically by Oersted and Ampere but its origin in a changing electric field was not understood until Maxwell gave the first full formulation of the electromagnetic field.

We can summarize their findings as follows. If we use as a detecting element a moving charge, such as the stream of electrons in a cathode ray oscilloscope, we find that forces are exerted on the electron beam which are not due to electrostatic fields. To explain these we postulate the existence of a *magnetic field* symbolized [2] by **H** which produces a force of **F** dynes where

$$\mathbf{F} = \frac{e\mathbf{v} \times \mathbf{H}}{c} \tag{1·2}$$

Here e is the magnitude of the moving charge in electrostatic units, $\mathbf{v}$ its velocity, and c a constant which is equal in magnitude to the velocity of light. The vector product $\mathbf{v} \times \mathbf{H}$ means that the direction of the force is in the direction of motion of a right-handed

[2] In this book we deal almost exclusively with free space where in this system of units **H** is identical with **B**. We have elected therefore to give this force equation in terms of **H** rather than **B** as is more usual.

screw when **v** is turned to the direction of **H**. The magnitude of the force is the product of the magnitude of ev and H times the sine of the angle between them.

This is a great deal more complicated than the equivalent relation $\mathbf{F} = e\mathbf{E}$, but, since it fits the experimental facts, it must be accepted.

Maxwell showed that the magnetic field as experimentally defined above is produced by a varying electric field. The relationship is inherently simple but involves a form of quantity which we are not accustomed to describe quantitatively. The magnetic field is produced by an electric field which is changing with time. The resulting magnetic field, however, is not directed along the direction of the electric field, nor is the field located at any point of space which is definite. The changing electric field produces a magnetic field which is spread out, but spread out in such a way that the field which is definitely related to the changing electric field is a distribution which contributes to a circular rotation. Such a spread-out rotationally controlled quantity requires an elaborate description, but it is exactly done by the quantity known in vector analysis as the *curl*. The relationship in terms of this is

$$\operatorname{curl} \mathbf{H} = \frac{1}{c}\frac{\partial \mathbf{E}}{\partial t} \tag{1·3a}$$

We can also write the relation in three-dimensional coordinates as

$$\begin{aligned} \frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z} &= \frac{1}{c}\frac{\partial E_x}{\partial t} \\ \frac{\partial H_x}{\partial z} - \frac{\partial H_z}{\partial x} &= \frac{1}{c}\frac{\partial E_y}{\partial t} \\ \frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} &= \frac{1}{c}\frac{\partial E_z}{\partial t} \end{aligned} \tag{1·3b}$$

This kind of relation is difficult to work with except for one feature. If a field such as **H** exists and threads through any closed loop we can relate the integral of the curl over the area w[illegible] integral of the magnetic field around the loop itself. N[illegible] we don't choose *any* closed loop; we choose a simple one in order to extract information from this relation with the minimum of

effort. Stokes' theorem expresses this relation, and in symbols it is

$$\oint \mathbf{H} \cos\theta \, dl = \int_{\text{surface}} \text{curl}\, \mathbf{H} \cos\phi \, dS$$

where the left-hand side refers to integration around the loop and θ is the angle between **H** and dl; and the right-hand side refers to integration over a surface bounded by the loop. Here ϕ is the angle between the normal to the increment of area, dS, and the direction of curl **H**. The meaning of curl and a discussion of Stokes' theorems are given in Appendix 2. Applying this to a loop in which the electric field is changing we get

$$\oint \mathbf{H} \cos\theta \, dl = \frac{1}{c} \int_{\text{surface}} \frac{\partial E}{\partial t} \cos\phi \, dS \qquad (1 \cdot 3c)$$

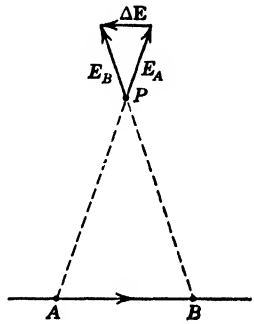


FIG. 1·1 Electric field near a moving charge. At a point P the field sweeps around from direction E_A to direction E_B. This requires a vector **ΔE** parallel to AB as indicated. It is this changing field which gives rise to the magnetic field around the direction AB.

The use of relation 1·3c can be seen by considering the magnetic field around an electron moving with uniform speed. If we consider a point a little to one side of the path of the electron the electric field which is along the line from the point to the electron sweeps around as the electron passes. dE/dt is therefore a vector which is easily constructed geometrically, as can be seen from Fig. 1·1, and is, for example, parallel to the direction of motion of the electron at the moment of closest approach. This gives rise to a magnetic field which is around the direction of motion of the electron; when the figures are put in it is seen to be of magnitude

$$H = \frac{ev \sin\theta}{cr^2} \qquad (1 \cdot 4)$$

where v is the speed of the electron, r the distance from the point to the electron, and θ the angle between the direction of r and that of the motion of the electron.

If this relation is applied to the motion of electrons through an element dl of wire and if there are N electrons per unit length of wire moving with an average velocity v, the number which pass a point per second is the number in a distance v, which is Nv. The current is therefore

$$i = Nev$$

The magnetic field (here represented by dH), however, is that due to the electrons, $N\,dl$ in number, moving in the element dl. In terms of current this number is $i\,dl/ev$. Then, using relation 1·4,

$$dH = \frac{1}{c}\frac{i\,dl \sin\theta}{r^2} \tag{1·5}$$

This is the familiar expression for *Ampere's law*. Notice that the most characteristic feature of a current, its magnetic field, can be thought of as a consequence of the changing electric field which is produced by the motion of charge. We wish again to stress the importance and reality of the effects in the space surrounding the moving charge.

Effect of a Changing Magnetic Field

The next question is whether the production of a magnetic field by a changing electric field is the final step. The answer is no. It was discovered by Faraday and Henry that, just as a changing electric field has a rotationally controlled magnetic field associated with it, a changing magnetic field has a rotationally controlled electric field. The familiar way of describing this is by the relation

$$\oint E \cos\theta\, dl = -\frac{1}{c}\int_{\text{surface}} \frac{\partial H}{\partial t}\cos\phi\, dS \tag{1·6}$$

which states that the electromotive force developed in a closed loop is equal to the negative time rate of the change of flux through the loop. Because this familiar relation is rather cumbersome for microwaves in many instances, for example in a plane electromagnetic wave, something more directly related to the fields is

more convenient. This is supplied by considering the above relation to be an integral form of the basic relation

$$\begin{aligned} \frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z} &= -\frac{1}{c}\frac{\partial H_x}{\partial t} \\ \frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x} &= -\frac{1}{c}\frac{\partial H_y}{\partial t} \\ \frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} &= -\frac{1}{c}\frac{\partial H_z}{\partial t} \end{aligned} \tag{1·7a}$$

or in its vector form

$$\text{curl}\,\mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{H}}{\partial t} \tag{1·7b}$$

In this version of Faraday's law, which is converted into the usual form given in equation 1·6 by the application of Stokes' theorem, the two fields are directly related. The form is rather complicated, but it can be greatly simplified in many cases, as shown in the next section.

Electromagnetic Waves

Before completing this brief account of the fundamentals of electromagnetic theory it is worth while to see one important result which can be derived from the two basic relations just given. These relations are between time derivatives and distance derivatives. The faster E is changing with time the more does H vary with distance. This is a characteristic necessary for the propagation of a wave. In an elastic solid the distortion present when a wave is traveling through provides a force which causes the acceleration of the medium, and in this way space and time derivatives are connected. The equation connecting them involves the elasticity and the density, and these control the velocity of propagation. Now suppose a very simple process to be set up, namely, the oscillation of an electric charge up and down. It is quite reasonable to suppose that this sets up an electric field E_y at a point distant z along the z axis at a time t, and it can also be assumed that we choose a direction such that the x and z components of E vanish. Later we show that this can be done. The

electric field is thus plane polarized. Having thus set up a very simple kind of field we can now inquire into the consequences of equations 1·3*a* and 1·7*a*. Substituting the above value for the electric field in these yields

$$-\frac{\partial E_y}{\partial z} = -\frac{1}{c}\frac{\partial H_x}{\partial t} \qquad \frac{\partial E_y}{\partial x} = -\frac{1}{c}\frac{\partial H_z}{\partial t}$$

$$\frac{\partial H_x}{\partial z} - \frac{\partial H_z}{\partial x} = \frac{1}{c}\frac{\partial E_y}{\partial t}$$

To reduce the problem to its simplest form, we further stipulate that there is no magnetic field along the z axis. Then there result the equations

$$-\frac{\partial E_y}{\partial z} = -\frac{1}{c}\frac{\partial H_x}{\partial t}$$

$$\frac{\partial H_x}{\partial z} = \frac{1}{c}\frac{\partial E_y}{\partial t}$$

Differentiating the first with respect to t, the second with respect to z, we get

$$-\frac{\partial^2 E_y}{\partial t\,\partial z} = -\frac{1}{c}\frac{\partial^2 H_x}{\partial t^2}$$

$$\frac{\partial^2 H_x}{\partial z^2} = \frac{1}{c}\frac{\partial^2 E_y}{\partial z\,\partial t}$$

from which is obtained

$$\frac{\partial^2 H_x}{\partial z^2} = \frac{1}{c^2}\frac{\partial^2 H_x}{\partial t^2}$$

A similar equation for E_y can be derived by eliminating H_x.

This equation represents a traveling wave. Its solution is of the form

$$\left.\begin{matrix} H_x \\ E_y \end{matrix}\right\} = f\left(t - \frac{z}{c}\right) \tag{1·8}$$

a special case of which is the familiar sine wave

$$\left.\begin{matrix} H_x \\ E_y \end{matrix}\right\} = A \sin \omega\left(t - \frac{z}{c}\right) \tag{1·9}$$

This represents a wave, of amplitude A, traveling along the z axis in the direction of z, increasing with a velocity c. The quantity ω controls the rate at which E_y changes with time and ω/c the rate at which E_y changes with distance. $\omega/2\pi$ is the frequency f, and $2\pi/(\omega/c)$ is the wavelength λ.

This far-reaching consequence of the nature of electric and magnetic fields was first pointed out by Maxwell. No physical interpretation of the wave motion in terms of forces and accelerations can be given, but it is agreeable and significant that the quantity c, which appears in both fundamental equations and which is therefore to be thought of as basic to the fields, is the velocity of propagation.

Notice that the electric and magnetic fields are equal in amplitude and are in phase. Both are at a maximum at once, and both vanish at once.

This particular case of an infinite plane-polarized wave (infinite because no boundary effects are treated) can rather easily be shown to be a simplification of a three-dimensional wave equation with a solution which is essentially the same. To show this compactly requires vector analysis: if the reader has patience he can work through the six equations of $1 \cdot 3b$ and $1 \cdot 7a$ by the same procedure and obtain the general result.

Energy in the Magnetic Field

In the same way that we supposed energy to reside in the electrostatic field, we suppose energy to be present in the magnetic field. By carefully accounting for the work required to establish a magnetic field by any process we can show that the energy in ergs per cubic centimeter is $H^2/8\pi$. This is seen to be symmetrical with the expression for the energy of the electric field.

Transfer of Energy, Poynting Vector

The total energy in an electromagnetic field is therefore

$$\frac{1}{8\pi} \int_{\text{volume}} (E^2 + H^2)\, dv$$

If this energy is the result of a process which makes the two fields interrelated, we can inquire about the time derivative of this expression. This derivative will tell us the flow of energy, which is of considerable interest.

If we consider the energy to be enclosed in a volume, take the time derivative, and substitute the appropriate values for $\partial\mathbf{E}/\partial t$, $\partial\mathbf{H}/\partial t$ from the fundamental equations $1 \cdot 3a$ and $1 \cdot 7b$, we find that the total loss of energy by this volume per second is

$$\frac{c}{4\pi} \int_{\text{surface}} \mathbf{E} \times \mathbf{H} \cdot d\mathbf{S} \qquad (1 \cdot 10a)$$

If we invent a quantity $\mathbf{P}$ called the Poynting vector, which is such that the surface integral of the outward component of this new vector is the same as the total loss of energy, we have

$$\mathbf{P} = \frac{c}{4\pi} (\mathbf{E} \times \mathbf{H}) \qquad (1 \cdot 10b)$$

The Poynting vector is a most interesting quantity. Applying it to the flow of current in a wire is very instructive. If the wire is without resistance the only electric field is radially outward and the only magnetic field is around the wire. The whole flow of energy is then along the direction of the wire. The energy is, however, in the field, not in the wire. If the wire has resistance there has to be a component of electric field along the wire or current will not flow. This component of field is perpendicular to the circular magnetic field, but the vector product of the two is a Poynting vector directed inward to the wire. This flow of energy out of the field into the wire is the necessary supply of energy to provide the increased thermal agitation of the molecules in the wire. The field as the location of the energy is again seen.

The Poynting vector for a plane wave oscillates in value but maintains the direction of the wave.

Production of Radiation: Field of an Accelerated Charge

On page 5 the factors on which the electric field are dependent have been listed. It was pointed out that an important factor is the time during which a charge causing the field has been at the

particular location. Today it is possible to create electric charge by the methods of nuclear physics. This does not produce a field at a distant star at once. The field is propagated with a finite velocity, the velocity of light. Hence, if we are interested in an electric field due to a charge at a distance r, we are not at the moment interested in the charge at that point now, but in the charge which was there a time r/c seconds earlier.

This simple and important fact was not discovered at once. It awaited a rather mature development of electrodynamics before it was clearly realized. It is still the custom to discuss electrodynamics in terms of fields which are maintained so long that this time dependence does not appear, and then to point out that Maxwell's equations (1·3 and 1·7) permit a general solution which involves the kind of retarded field just described. It can, nevertheless, be considered a fundamental property of the electromagnetic field, and for the purpose of this brief review we adopt that approach.

Since it is necessary to state clearly how the electric field is developed around a charge in order to see how radiation is developed, we describe it as follows. The charge can be considered to be continually emitting a field. This is essentially what we mean by electrification. The field is emitted with a finite velocity, denoted by c, which is also the velocity of light. The field is emitted in all directions equally for charges which move slowly and uniformly. *During such motion the field distribution as seen from the charge remains the same.*

If, as we have just asserted, the field distribution as seen from a moving charge with fixed velocity remains the same, there now is the question about the distribution when the charge changes velocity. This *cannot* remain the same, and to see what happens we can imagine the following process. Two equal charges are placed at O (Fig. 1·2) for a considerable time. Then one of them is given an acceleration a for a short time t, which gets it to O', and it thereafter coasts with uniform velocity v. Considering one direction of emission, say at angle θ, the fixed charge produces a field represented by the line $OCC'C''$. The moving charge, after it begins to coast, produces the same field, at the angle θ, but this field, represented by AB or $A'B'$, has only been established for a short time, namely the time T to travel from O' to A, in the first case, or T' to A' in the second. The field therefore endures only

for a finite distance AB or $A'B'$ and then becomes identical with that of the equal stationary charge. We are thus led to assert that an accelerated charge emits a field which is *not along one direction, but has a kink at one localized region,* as denoted by BC, $B'C'$, or $B''C''$ in the figure.

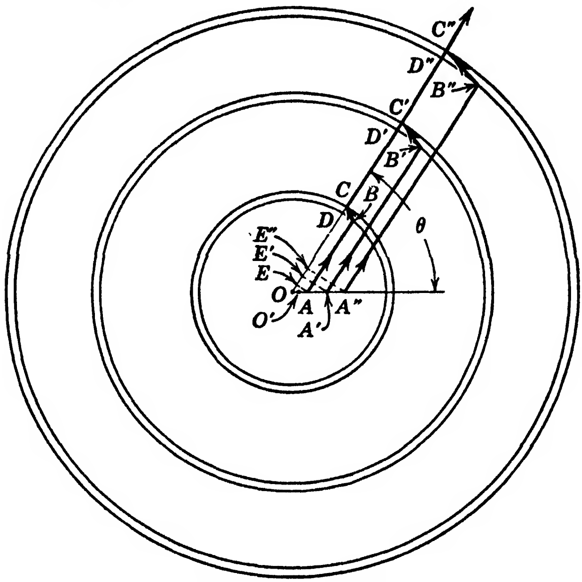


FIG. 1·2 A method of visualizing radiation by an accelerated charge. Since the electric field travels with a finite velocity, there must exist a region where the transition from the new field of a moving charge joins the old field which existed when it was stationary. This transition region is as drawn and plainly has a component perpendicular to the line from the original position of the charge. This method of visualization is due to Stokes, who used it to explain the formation of x-rays.

If this assertion is correct it should be possible to apply some analysis to the process. Thus we note that, as the charge coasts farther, the kink becomes greater, and also farther away. Using a triangle like BDC we see that the ratio of the purely transverse component BD to the radial component CD increases. This ratio has the value CD/EA or $C'D'/E'A'$, and so on, and, since $CD = C'D' = C''D''$ and $EA = OA \sin\theta$, $E'A' = OA' \sin\theta$, we can obtain an explicit value for the transverse field providing we have expressions for CD and OA, and so on. These are $CD = ct$, OA

$= vT$ approximately $= atT$, $OA' = vT' = atT'$. Now the interest really lies in the field at various distances from the charge, rather than in terms of T, T', and so on. If we denote AB by r, then T is the time for the field to travel the distance r, which is r/c. Then $OA = atr/c$, and so on.

The ratio of the transverse field E_t to the radial field (which is e/r^2) is

$$\frac{E_t}{\frac{e}{r^2}} = \frac{EA}{CD} = \frac{\frac{atr \sin \theta}{c}}{ct}$$

or

$$E_t = \frac{ea}{rc^2} \sin \theta \tag{1·11}$$

This expression is exceedingly interesting. The radiation field depends on the acceleration of the electron, the sine of the angle between the direction of acceleration and the radius considered, and the inverse first power of the radius.

It can be shown that an identical expression holds for the magnetic field; therefore the magnetic field in gauss is equal to the transverse electric field in statvolts per centimeter.

Energy Flow

The remarkable fact that a transverse electric field exists, together with a magnetic field of equal magnitude perpendicular to it, both being perpendicular to the radius vector, enables us to use the Poynting vector to determine the flow of power. We have already done this for an electron which is not accelerated, the equivalent of a current in a wire with no resistance, and have shown that the flow of power is along the direction of motion of the electron. In the case now considered the power radiated by the radiation field is

$$\frac{c}{4\pi} \int \mathbf{E} \times \mathbf{H} \cdot d\mathbf{S}$$

Substituting the values from equation 1·11 gives

$$\frac{c}{4\pi} \int \frac{e^2 a^2}{c^4 r^2} \sin^2 \theta \, dS$$

Integrated over a sphere this gives for the total energy radiated per second

$$\frac{2}{3}\frac{e^2a^2}{c^3}$$

Notice that the maximum rate of radiation which occurs when θ is 90°, or perpendicular to the acceleration, is given by

$$\frac{e^2a^2}{4\pi c^3 r^2}\,dS$$

If power were radiated uniformly at this rate the total would be

$$\frac{e^2a^2}{2c^3}\int_0^{\pi}\sin\theta\,d\theta = \frac{e^2a^2}{c^3}$$

The ratio of this to the actual value is $3/2$. This ratio is a particular case of an "antenna gain" (cf. page 124). In this case the antenna is a single accelerated electron.

The electric field of radiation is always perpendicular to the radius vector. It always has a component in the direction of the electron acceleration except for the two directions where the radiation field is zero. The magnetic field is perpendicular to the electric field and changes magnitude with it. A plane wave such as was considered on page 12 can therefore be realized by considering the field a long way from an accelerated electron.

A slightly more complicated case is the field near an oscillating dipole. We now have two added factors. The first is a smoothly varying value of a, the acceleration of the charge; the second is the presence of both signs of charge. Neither of these proves difficult, and we can actually use the above expressions if we substitute the correct value for the acceleration which is determined by the amplitude and frequency.

Thus if $z = A \sin \omega t$ represents the oscillation of an electron in the dipole, $d^2z/dt^2 = -A\omega^2 \sin \omega t$. This is substituted for a in expression $1 \cdot 11$.

Fields near a Moving Charge

The three major fields near a moving charge are therefore

1. e/r^2, the static Coulomb field—radial
2. $(ev \sin \theta)/cr^2$, a magnetic field due to the moving charge
3. $(ea \sin \theta)/c^2r$, electric and magnetic radiation fields.

There are other fields due to the interplay between changing electric and magnetic fields. Thus, when a magnetic field corresponding to 2 is produced, the process of formation of the field implies a time derivative of a magnetic field which gives rise to an electric field. Such a field is called an induction field. Only the radiation fields are important at distances exceeding a few wavelengths because they vary more slowly with r.

Influence of the Medium

It will be noticed that all considerations have applied to empty space. For the majority of applications this is sufficient. Nevertheless it is necessary to recognize that electrical and magnetic effects take place in solids, liquids, and gases, and that the nature of the medium plays a part.

Electric, magnetic, and other fields are paramount in determining the structure of matter. Therefore there is bound to be an interaction between material substances and an impressed field. Thus, in the case of a ferromagnetic body a rearrangement may call into being a new field which is far greater than the original causing agent. Or, in the case of a conductor, the effect of the field may be to cause a redistribution of charge which will eliminate the field altogether.

Two strong effects of matter exist. The first is the possibility in certain substances, notably metals and electrolytes, that electric charges, either electrons themselves or ions, are free to move. This freedom is, however, hampered by the presence of enormous numbers of neutral molecules, through which the electrons have to drift. Drifting consists of brief accelerated passages, in which the influence of the external field is felt, followed by collisions in which the atomic field exerts a far greater influence and sends the electron or ion on a totally different path. The general effects of collisions are random, whereas the influence of the field is directed. Hence an ion actually does progress in the direction of the field. The net result is that an average velocity v can be ascribed to an electron or ion. This average velocity is proportional to the applied electric field; therefore the relation

$$v = bE$$

where b is a constant depending on the substance, is obtained.

The current flowing across an area A is the number of electrons crossing that area per second. This is N, the number per cubic centimeter, times the volume of a cylinder of length defined by the distance gone in one second, namely v, and area A, which is NAv. Using this with the relation above gives for the current i

$$i = NAbEe$$

If a quantity V, the potential difference, is defined as the work done in taking a unit charge through a length l of the material through which the electrons are moving, which is generally a wire, against the field E,

$$V = Eel$$

so that

$$i = \frac{NbAV}{l}$$

or

$$V = \left(\frac{l}{A}\right)\left(\frac{1}{Nb}\right)i \qquad (1 \cdot 12)$$

This is *Ohm's law*. The quantity $1/Nb$ is called the *resistivity*, and Nb itself is called the *conductivity*. The higher the number of free electrons, the greater is the conductivity. The higher the temperature is, the greater the random motion and the lower the value of b; therefore the conductivity is lower.

This first effect exists in matter where it is possible for a large scale motion of electrons to occur. In a wide class of substances this is not possible, all electrons being tightly bound to their appropriate atoms. This means that no large scale transport of charge can ever exist. The electric and magnetic fields applied to this kind of matter are not without effect in spite of this, since they can cause rotation of molecules or the displacement of the positive and negative charges in an atom or molecule with respect to one another. The total effect of the applied field and the distorted atoms will certainly be different from the original field. In some cases it is possible to estimate what this effect will be, as for example in a noble gas. In others the general nature of the effect can be seen after it has been studied, but so far no sure predictions can be made. In this class lies ferromagnetism.

As generally happens when underlying phenomena are complicated, the method of handling the theory is subject to some

variation. Most commonly, it is stated that in the presence of electric and magnetic material (that is, anything) two new fields exist which are defined by

$$\mathbf{D} = K\mathbf{E}$$

$$\mathbf{B} = \mu\mathbf{H}$$

The first is an electric field which is related to the field which would be present in free space, the second a magnetic field similarly related. The quantities K and μ, the dielectric constant and permeability, are not fixed constants. They depend on the material and also on the field. Actually K and μ can be complex quantities, another way of saying that **D** and **E** or **B** and **H** are not in phase with one another. Often K and μ are capable of interpretation in terms of a mechanism by which the field is formed, for example the distortion of atomic structure to form dipoles, or the rotation of atomic dipoles into a more orderly pattern than thermal agitation will permit in the absence of an applied field. The attitude taken here is that K and μ can be experimentally measured in many cases and are often constant over a moderate range of fields. In our system of units K and μ are both unity for free space and nearly so for air.

The importance of these composite fields lies in the fact that the distance variations of **H** and **E** are caused by the time variation of **D** and **B**. That is, the two fundamental relations $1 \cdot 3a$ and $1 \cdot 7b$ become

$$\operatorname{curl} \mathbf{H} = \frac{1}{c}\frac{\partial \mathbf{D}}{\partial t}$$

$$\operatorname{curl} \mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{B}}{\partial t}$$

This statement means that a primary field such as **H**, which is varying, pulls with it atomic fields in the material, so as to produce a total field **B**. The time variation of **B** is now the quantity which determines the induced electric field by Faraday's law. A similar statement holds for the induced magnetic field due to a varying electric field.

In general it is desirable to set up these fundamental equations, generally called Maxwell's equations, so that they require the minimum of reformulation before use. It is found convenient to

treat an actual current, where moving charge is concerned, as a separate causative agent of magnetic field because in so many problems currents can be seen as separate entities. (Actually, of course, the varying electric field produced by the current produces the magnetic field.) We give here the form of Maxwell's equations in which current is included explicitly. **u** is the current *density* in statcoulombs per square centimeter per second:

$$\text{curl}\,\mathbf{H} = \frac{1}{c}\left(4\pi\mathbf{u} + \frac{\partial\mathbf{D}}{\partial t}\right)$$

or

$$\begin{aligned}
\frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z} &= \frac{1}{c}\left(4\pi u_x + \frac{\partial D_x}{\partial t}\right)\\
\frac{\partial H_x}{\partial z} - \frac{\partial H_z}{\partial x} &= \frac{1}{c}\left(4\pi u_y + \frac{\partial D_y}{\partial t}\right)\\
\frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} &= \frac{1}{c}\left(4\pi u_z + \frac{\partial D_z}{\partial t}\right)
\end{aligned} \qquad (1\cdot 13a)$$

$$\text{curl}\,\mathbf{E} = -\frac{1}{c}\frac{\partial\mathbf{B}}{\partial t}$$

or

$$\begin{aligned}
\frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z} &= -\frac{1}{c}\frac{\partial B_x}{\partial t}\\
\frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x} &= -\frac{1}{c}\frac{\partial B_y}{\partial t}\\
\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} &= -\frac{1}{c}\frac{\partial B_z}{\partial t}
\end{aligned} \qquad (1\cdot 13b)$$

where the subscripts refer to the three components.

In addition to these equations three others are available. The first is a statement of Coulomb's law in some form. The second is a statement that free magnetic poles are not found. The third is Ohm's law. For the purposes of calculation where field equations are needed Ohm's law is generally cast in a different form. Writing it as

$$V = \gamma\frac{i}{A}l \qquad (1\cdot 14)$$

where V is the applied potential difference, γ the resistivity, i the total current crossing an area A, and l the length of the conductor,

we can put $V = El$. Then

$$E = \gamma u$$

where u is the current density. For three dimensions this is written vectorially as

$$\mathbf{E} = \gamma \mathbf{u}$$

In this equation $\mathbf{E}$ is in statvolts per centimeter and $\mathbf{u}$ is in statamperes per square centimeter. Making the conversion of units, γ is the resistivity in ohm-centimeters divided by 9×10^{11}. It is therefore a very small quantity for a conductor compared to K and μ which are near unity. This feature dominates the propagation in a conductor.

This concludes the review of the basic equations of electrodynamics. We can add one more remark. A great part of useful electrical thinking is in terms of Ohm's law which is stated in terms of a quantity of the form $\int E\,dl$. The field equations, when cast in the integral form by Stokes' theorem, lend themselves to this process. For instance equation 1·6 already contains $\int E\,dl$, and the right-hand side of this equation is the basis for definition of *inductance*. Therefore the more usual circuit equations are fundamentally derived from those given here. This is convenient as long as the fields are separated into localized places, as in coils of wire and between the plates of condensers. It ceases to be convenient when both fields are distributed throughout the apparatus and is exceedingly cumbersome to apply to free space radiation. In the microwave art we find we need both forms of thinking. We have presented here the less familiar description of electrical behavior to enable the use of both.

With the equations of this section we should be able to work out any theoretical problem involving microwaves and phenomena not involving transitions within atoms (where the above theory has to be modified). This we can now follow up.

Propagation of a Plane Wave in a Conductor: Skin Depth

The essential feature of microwave technique is the development of field energy in a space bounded by conductors. This field energy is then used for various purposes, such as accelerating

electrons for their own sake or for the secondary purpose of maintaining oscillations, and so on. In any event, it is of paramount importance to know what happens to an electromagnetic wave in a conductor, and this is turn permits the answering of the important question of what the conductor does to the wave outside it. These two subjects, which are generally referred to as "skin depth" and "boundary conditions" can now be considered. The first can be illustrated by the simplest case of wave propagation in a conductor—the propagation of a plane wave.

For the purpose of this problem and many other purposes wave processes are represented by means of complex exponentials. The great value of these is their relation to a rotating vector which is of great help in describing oscillatory phenomena. Thus when a wave is written as a complex exponential a quantity is used which can readily be analyzed into real and imaginary parts. The real part is a cosine, the imaginary part is a sine multiplied by $\sqrt{-1}$, written j. Choosing the operation of multiplying by j to denote rotation through a right angle leads to no analytical contradictions; therefore the imaginary part represents the y component of the vector, the real part the x component, and the quantity itself is a direct analytical representation of the generating vector. Since in most processes involving summation of waves the components of the *vector* sum of the generating vectors are required for an explicit statement of field voltage or current, the ability to add these vectors analytically is a great convenience.

In bare essentials a rotating vector is described by a quantity $Ae^{j\theta}$ where A is an amplitude and θ an angle called the phase. In a wave, however, the angle is a function of time and distance. If the relation to time is denoted in terms of an angular velocity ω, and to distance in terms of a constant k, then, for a wave traveling in the forward direction, there is a diminution of the phase angle at larger values of x. Hence

$$\theta = \omega t - kx$$

Now if this is to represent a wave of frequency f, wavelength λ, and velocity c, ω and k can be expressed as follows. The angle traversed in 1 second is ω. It is also f revolutions or $2\pi f$ radians. Hence $\omega = 2\pi f$. Also λ denotes the fact that in traveling a wavelength the phase changes by 2π. The change of phase per centimeter is therefore $2\pi/\lambda$. This is the same as k. Hence $k = 2\pi/\lambda$.

In addition, if the wave advances an amount λ, f times per second, the velocity is $f\lambda$. Hence $c = f\lambda$.

The three relations

$$\omega = 2\pi f$$

$$k = \frac{2\pi}{\lambda}$$

$$c = f\lambda$$

permit several variations of the explicit form for a wave to be written; for example,

$$E_y = Ae^{j\omega[t-(x/c)]} = Ae^{j2\pi[ft-(x/\lambda)]}$$

Notice that if the wave travels in a medium of refractive index n the velocity is c/n, and the first form of the equation becomes

$$E_y = Ae^{j\omega[t-(nx/c)]}$$

The propagation of a plane wave such as the above in a metal can now be examined. Suppose

$$E_y = Ae^{j\omega[t-(nx/c)]}$$

$$H_z = Ce^{j\omega[t-(nx/c)]}$$

$$B_z = \mu H_z$$

$$D_y = KE_y$$

$$E_y = \gamma u_y$$

Substituting these expressions in equations $1\cdot 13a$ and $1\cdot 13b$, which is remarkably simple as each set of six terms reduces to one only, there result

$$Cj\omega n = \frac{4\pi A}{\gamma} + j\omega KA$$

$$An = \mu C$$

which reduce to the single equation

$$n^2 = K\mu - j\left(\frac{4\pi\mu}{\omega\gamma}\right) \qquad (1\cdot 15)$$

The quantity n is therefore partly imaginary. Writing it as $n = p - jq$, there results

$$E_y = Ae^{-\omega qx/c}e^{j\omega[t-(px/c)]} \qquad (1\cdot 16)$$

The first exponential indicates an amplitude which diminishes as x increases. When x has the value $c/q\omega$ the real exponent is -1 and the amplitude is down to $1/e$ or 37 per cent. This value of x is the *skin depth.* Giving its value is one way to express the way in which the plane wave penetrates into the conductor. In order to determine the value of the skin depth, the value of q in terms of quantities appropriate to the conductor must be known. Substituting $n = p - jq$ in equation $1 \cdot 15$ there result

$$p^2 - q^2 = K\mu$$

$$pq = \frac{2\pi\mu}{\gamma\omega}$$

Now it has already been remarked that γ is small compared to K and μ for a conductor, so that the term containing $1/\gamma$ is by far the largest, or in other words $p^2 - q^2$ can be neglected in comparison with pq—another way of saying that to a close approximation $p = q$. Then

$$q = \sqrt{\frac{2\pi\mu}{\gamma\omega}}$$

or, because $\omega = 2\pi c/\lambda$,

$$q = \sqrt{\frac{\mu\lambda}{\gamma c}}$$

The skin depth is then

$$\frac{1}{2\pi}\sqrt{\frac{\gamma\lambda c}{\mu}}$$

Table $1 \cdot 1$ gives some values for the skin depth for several conductors and several wavelengths.

TABLE 1·1

Conductor	Skin Depth (Centimeters)			
	$\lambda = 1$ cm	$\lambda = 10$ cm	$\lambda = 100$ cm	$\lambda = 5 \times 10^8$ cm (60 cycles)
Copper 20°C	3.78×10^{-5}	1.19×10^{-4}	3.78×10^{-4}	0.85
Copper −258°C	3.45×10^{-6}	1.09×10^{-5}	3.45×10^{-5}	0.077
Silver 0°C	3.52×10^{-5}	1.11×10^{-5}	3.52×10^{-4}	0.79
Gold 20°C	4.53×10^{-5}	1.43×10^{-5}	4.53×10^{-4}	1.02
Lead 20°C	1.39×10^{-4}	4.40×10^{-5}	1.39×10^{-3}	3.12
Brass 20°C	7.7×10^{-5}	2.4×10^{-4}	7.7×10^{-4}	1.73

A little further analysis shows that almost all the energy inside a conductor is magnetic. This is not surprising because the chief feature of a conductor is the ability to develop high currents for low fields. The high currents carry the energy and it is magnetic.

The skin depth is a most important quantity for it determines the selectivity of resonators and the losses of transmission lines.

Boundary Conditions

The two simple cases, transmission in free space and transmission in a good conductor, have just been considered. Now arises the important question as to what occurs when an electric field and a magnetic field are established partly in the free space and partly in a conductor. In most treatments of electricity these boundary conditions are discussed briefly and usually prove to be of little further interest because the actual fields are established in coiled wires and flat plates. Microwave fields are different: they are established almost always in cavities or in conducting tubes, and the somewhat academic question of boundary effects becomes of central importance.

Boundary conditions are, of course, determined by the behavior of the conductor at its surface. Three things can clearly be said about the conductor: it can develop a surface charge; surface currents can flow; and no fields, electric or magnetic, will be found inside beyond about a millimeter at most. The question is therefore how these three facts must influence the field outside the conductor. A complete treatment can readily be given by using the full electromagnetic equations, and these should be consulted by anyone interested in following the subject deeply. A great deal of physical insight can be obtained from Fig. 1·3. In Fig. 1·3 (*a*) is shown a single upper layer of positive charge, produced in the conductor by the influence of the applied field. The arrows indicate the direction of the force from each charge. It can be seen that the arrows tangential to the surface cancel one another, while the arrows normal to the surface add up. In other words the effect of a surface layer of charge is the development outside the conductor of a *normal field*, but not a tangential field. In Fig. 1·3 (*b*) is shown a similar diagram for surface *currents*. The magnetic field around each current element is shown, and it can

be seen that the different character of a magnetic field has a clear-cut consequence for now there is no net normal field but a net *tangential field.* Therefore a combined surface charge and surface current distribution enables the surface to produce a normal electric field and a tangential magnetic field. The surface will not

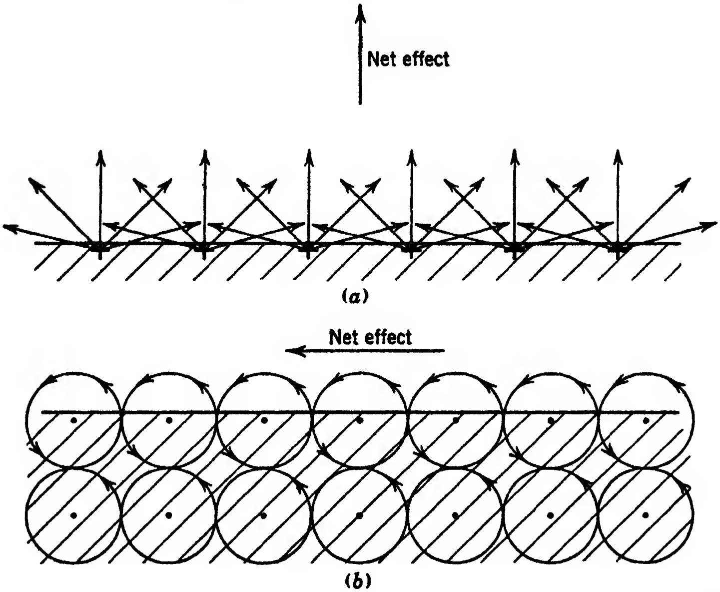


FIG. 1·3 Boundary conditions: (*a*) shows that the net field due to a surface layer of static charge is a normal component of electric field; (*b*) shows that a surface layer of currents (represented by dots indicating currents flowing out of the paper) gives a net magnetic field which is tangential.

exert any effect on a normal magnetic field or a tangential electric field.

The considerations just outlined hold for any surface. When the surface is a conductor there is the added fact that within a very small distance inside all fields vanish. This exerts no great limitation on the normal electric field, for we can suppose that a large enough charge is pulled to the surface to produce the necessary sudden change to zero. In the same way there is no limitation on the tangential magnetic field, for surface currents can be supposed to produce the sudden change of magnetic field. On the other hand the surface is no help at all with regard to a tangential elec-

tric field or a normal magnetic field. Since they are zero inside, it is to be expected that they will also be zero outside, at least in any condition which is steadily and stably maintained.

We therefore have the following concept of an electromagnetic field arriving at the surface of a conductor. Surface charges and currents are quickly established so that the normal component of E and tangential component of H can discontinuously (or almost so) drop to zero inside. The tangential E and normal H represent a strongly anomalous condition which will result in a rapid absorption of that part of the field. Now if the electromagnetic field is confined between several conductors and a short time allowed for a steady state to establish itself, the steady state will have no tangential E or normal H at any surface. Notice that there is no inherent restriction on these fields; the restriction holds only when a steady equilibrium state is established. More elaborate calculations show that the steady state, in the sense of having no tangential E or normal H, is established in a few cycles at most. Therefore, for any steadily established situation, the boundary requirements are that at the surface of a conductor the electric field will be accurately normal to, and the magnetic field will be parallel to, the surface.

It will be seen in the next chapter that these requirements limit the wavelength of the field in a conducting enclosure to a set of discrete values.

Reflection of Plane Waves

The above considerations can readily be applied to the reflection of plane waves by a conductor. Suppose this is at right angles to the paper along the left-hand edge. Consider perpendicular incidence as the simplest case. There is no normal component of the electric field, but there is a tangential component. There is also a tangential magnetic field but no normal field. The boundary conditions require that in the steady state there is no tangential electric field, but there can be a tangential magnetic field. This is achieved directly by a reflected wave (that is, one traveling in the opposite direction) of equal amplitude. Suppose that the incident wave, going from right to left, has the electric field up and the magnetic field into the paper. The reflected wave must, by the boundary conditions, have the electric field down, and since it is travel-

ing from left to right the magnetic field must also be into the paper. There is, therefore, double the magnetic field at the surface, and no electric field.

This is quite reasonable. The magnetic field is doubled *outside* and cancelled *inside* by strong surface currents. The currents reradiate energy in all directions. Outside it takes the form of a reflected wave; inside it exactly cancels the incident wave.

Standing Waves

The effect of reflection such as the above is to send a return wave through the space outside the metal at the same time as the incident wave is still arriving. This is a well-known physical phenomenon, and it yields standing waves. Consider for example the value of the magnetic field which results from the two waves. It is given by

$$H_{\text{incident}} + H_{\text{reflected}} = A\left(e^{j2\pi[ft-(x/\lambda)]} + e^{j2\pi[ft+(x/\lambda)]}\right)$$

which with a little rearrangement gives

$$H_{\text{total}} = 2Ae^{j2\pi ft}\cos\frac{2\pi x}{\lambda} \tag{1·17}$$

This is seen to be a region of space in which the field is varying with time with the original frequency but with an amplitude which is fixed at any point in space and varies according to the cosine of x. It will be seen later that the study of standing waves is most important in microwave technique.

We can now proceed to apply these fundamental ideas to the particular problems which arise in the production and transmission of microwaves.

Summary

In this chapter a brief review of electromagnetic theory from the point of view of fields has been given. The basic equations have been presented in a form suitable for considering the way in which fields are propagated in free space and in conductors. The penetration of fields into conductors has been considered and an expression for the skin depth deduced. The modifying action of conductors on the free space field, the so-called boundary condi-

tions, have been developed and applied to the case of the reflection of a plane wave.

REFERENCES

The material of this chapter owes a great deal to the following texts:

M. Abraham and R. Becker, *Classical Electricity and Magnetism*, translated by John Dougall, London, Blackie & Son, Ltd., 1932.

L. Page and N. I. Adams, *Electrodynamics*, D. Van Nostrand Co., 1937.

J. H. Jeans, *Electricity and Magnetism*, Cambridge University Press, 1909.

H. H. Skilling, *Fundamentals of Electric Waves*, John Wiley and Sons, Inc., 1942.

J. H. Van Vleck, *Electric and Magnetic Susceptibilities*, Oxford University Press, 1932. (Contains a most informative short account of the meaning of **E**, **D**, **B**, and **H**.)

PROBLEMS

1·1 Plot a graph showing the energy density in the space between two cylinders of radii 0.1 cm and 1 cm with a potential difference of 10,000 volts between them.

1·2 Calculate the energy in the gap of a cyclotron magnet 60 in. in diameter, 10 in. deep, with a field of 16,000 gauss.

1·3 Plot the radiation field of an electron subject to an accelerating field of 300 volts per centimeter for a microsecond.

1·4 Derive Ampere's law for the motion of a single charge with velocity V. Treat only the case for which $\sin\theta = 1$.

1·5 Compare the effectiveness of iron and copper sheets 1 to 2 mm. thick as shields against 10 megacycles per second and 100 kilocycles per second radiation.

1·6 Carry through the integration of the Poynting vector over a sphere and show that the value of the total energy radiated per second is $\frac{2}{3}(e^2a^2/c^3)$.

1·7 Apply Stokes' theorem to equation 1·13*a* for the case where only current flows and prove that $\oint H\,dl = (4\pi/c)$ (total current enclosed).

1·8 Apply the above relation to prove that the peak current i_1 across unit length of a conducting surface is $i_1 = H/2\pi c$, where H is the magnetic field in a normally incident plane wave.

CHAPTER 2

COAXIAL LINES, WAVEGUIDES, AND CAVITIES

Microwave techniques are largely concerned with radiation which is confined in some way. This is done either for the purpose of transmitting power where it is needed, or else for building up a large field, which is generally then used to influence an electron stream and so maintain oscillations. In this chapter we accordingly consider coaxial line and waveguide, which come under the first class, and resonant cavities, which are in the second.

The approach to all three of these is essentially the same. The boundary conditions will require that the electric field be normal to a conductor, the magnetic field tangential to it. These conditions, together with the basic electromagnetic equations, will determine the actual values of the fields. Naturally, very complicated solutions are possible, but, equally naturally, some very simple field configurations can be produced. By care in design the simplest configurations can be realized so that only rarely do complicated modes of electrical oscillation need to be considered. One of the simplest forms of oscillation is that in a coaxial line which is accordingly considered first.

2·1 COAXIAL LINES

For a great many purposes a coaxial line is an ideal method of power transmission. This is less true at microwaves than at wavelengths in the neighborhood of 50 centimeters, but the fact remains

that an understanding of coaxial lines is one of the greatest aids to the understanding of microwave technique. Also, the field configuration in a coaxial line is simple and easily understood. Therefore we consider coaxial line as the first method of power transfer.

Fields in a Coaxial Line

A coaxial line is a very special form of confined space. It has an outer cylinder which, as we shall see later, is able to determine

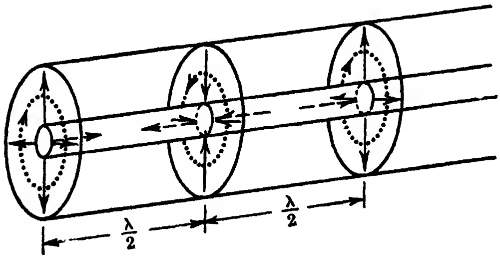


FIG. 2·1 Electric (solid arrows) and magnetic (dotted arrows) fields in a coaxial line. The electric field is radial, and the magnetic field is circular. Direction of propagation is from left to right. The current flow in the center conductor is indicated by dashed arrows.

to a very great extent the form of electric and magnetic fields which can pass down it. At the same time it has an inner cylinder which also has a strong effect on the field distribution. Therefore, as one might predict, the type of field found in a coaxial line is limited. Unless the wavelength of the fields is less than the radius of the outer cylinder only one mode can be transmitted. This is the mode which has the electric field radial and the magnetic field circularly around the axis, so fulfilling the boundary conditions. A diagram of this mode is shown in Fig. 2·1. The strong limitation imposed by the two conductors is mitigated by another consideration, namely the fact that both inner and outer conductors can supply charge and current distributions which make possible any rate of variation of the electric and magnetic fields along the direction of travel. For this reason the coaxial line is not frequency-sensitive. Whatever frequency may be applied to the line, the currents adjust themselves so that the field is transmitted. One other point can be made, namely that the electric and mag-

netic vectors are exclusively perpendicular to the direction of motion so that the velocity of propagation is the same as for a wave in the insulating medium without any conductors. The current flow is also indicated in Fig. 2·1. It oscillates back and forth as one would expect.

Since the amplitudes of the electric and magnetic fields do not fall off appreciably in value along the line it would be expected intuitively that they fall off as $1/r$, radially, this being a compromise between $1/r^2$, which holds for a single charge, and a constant value. This can be checked by an application of Maxwell's equations[1] to give, for the electric field, in statvolts per centimeter, and, for the magnetic field, in gauss:

$$E_r = \frac{A}{r} e^{2\pi j[ft-(x/\lambda)]}$$
$$H_r = \frac{A}{r} e^{2\pi j[ft-(x/\lambda)]} \tag{2·1}$$

where r is the radius at which the field is observed and A is an arbitrary amplitude. Now if we pick the point at which electric field, magnetic field, and current are all maximum, thus removing the need to consider the exponentials, we can see an interesting relation. Suppose the radius of the inner conductor is a and that of the outer b. Let $A = 1$, so that the electric field at the surface of the inner conductor will be $1/a$ statvolt per centimeter. Then the line integral of the electric field from center to outer conductor is $\int_a^b \frac{dr}{r} = \ln\left(\frac{b}{a}\right)$. This is the potential difference in statvolts across the cable. The magnetic field at a radius r is $1/r$ gauss. By applying Stokes' theorem to equation 1·13a as in Prob. 1·7, the current is found to be $4\pi i/c = (1/r)2\pi r$, where i is in statamperes. The current is therefore ½ abampere. The same total current flows in the outer conductor in the opposite direction. We therefore see that, for a given pair of radii, for the inner and outer conductors we have a calculable voltage difference, yet have the *same* currents flowing. The current is constant because it is determined by the line integral of magnetic field which in turn depends on the field at unity radius, which we fixed at unity. The

[1] See E. U. Condon, *Revs. Modern Phys.*, **14**, 357 (1942).

voltage is not constant because it depends on the distance we go in the electric field (that is, on the radii of the conductors). We therefore have the following result for the ratio of maximum voltage across the coaxial line to maximum current:

$$\text{Ratio} = \frac{\ln \frac{b}{a}}{\frac{1}{2}} \text{ statvolts/abamperes}$$

Since 1 statvolt = 300 volts and 1 abampere = 10 amperes this becomes

$$Z_0 = 60 \ln \frac{b}{a} = 138 \log_{10} \frac{b}{a} \text{ ohms} \tag{2·2}$$

The quantity Z_0 is called the characteristic impedance of the line. The points of interest are that the voltage and the current are in phase, and that the impedance depends only on the ratio of the radii and is a real number.

Before continuing the discussion of coaxial lines we wish to summarize the results of reasoning in terms of fields as given above. The boundary conditions limit the type of wave transmitted to fields of the type given by equation 2·1. The velocity of propagation is the same as for unconfined waves in the same medium. For an amplitude A, a current flows which has the maximum value $A/2$ abamperes regardless of conductor radii. The voltage across the line does depend on the radii, and the ratio of voltage to current is $60 \ln (b/a)$ ohms.

Practical Use of Coaxial Cables: Reflections

Cable is always used for a purpose; therefore the results given above for a uniform, infinite cable will not be directly applicable in practice. The question arises, how does the field adapt itself to the existence of bends, supports, side branches, and input and output equipment? Skill in understanding the adaptation and turning it to use is what makes a complete microwave engineer.

In reality the adaptation is simple. The cable can transmit equally well in both directions. The introduction of some change in the cable therefore changes boundary conditions in some way,

and in order to keep the requisite continuities a reflected wave is set up. This wave can have arbitrary amplitude and arbitrary phase. Use of the variability of these two enables the new conditions to be met without violating the all-important field equations.

The basic study is therefore that of the summation of incident and reflected waves, already described as *standing waves*. We consider these more fully in Chapter 4.

2·2 WAVEGUIDES

After what has been said about coaxial line the question arises as to whether the center conductor is needed and whether a single long metallic pipe can be used to confine electric and magnetic fields. The first speculation about this was made by Lord Rayleigh in 1898.[2] His analysis of the fields inside rectangular and cylindrical hollow pipes showed clearly that such pipes could be used to transmit electromagnetic radiation. He deduced correctly the facts about the minimum frequency which could be transmitted down a pipe of given size. The treatment he gives is simple and is, like everything Rayleigh wrote, well worth reading. From the entire absence of comment, however, it can be inferred that Rayleigh regarded his solution more in the light of an interesting mathematical problem than of one having any physical, let alone engineering, interest.

Waveguides are nowadays out of the stage of study for their own sake. They are used for a purpose. This purpose is nearly always the transmission of power either for subsequent radiation into space or for laboratory use. Therefore waveguides which supply this power in a form which can readily be fed into an antenna or another line are most commonly found. By far the greatest footage of waveguide is rectangular guide with the width about double the depth. This size is chosen because it transmits only one mode over a fairly wide frequency range, has a definite direction of polarization and low loss compared to coaxial line, and is reasonably easy to manufacture. Accordingly we can consider such a rectangular waveguide first.

To find expressions for the fields in waveguide we consider the space inside the guide. The fields there must obey Maxwell's

[2] Lord Rayleigh, *Phil. Mag.*, **43**, 125 (1898).

equations, and in addition the boundary conditions imposed by the presence of the rectangular conductor must be respected. The usual procedure is to obtain solutions for Maxwell's equations for various forms of propagation and apply the boundary conditions afterward. To do this in general is really simple, but tedious, as it involves the multiple equations of three dimensions with two kinds of fields. We can show the principle fully, without much tedium, by selecting a very simple form of transmission. The more general case is worked out clearly in Slater's *Microwave Transmission*,[3] to which we refer the reader for a more complete treatment.

We suppose that, with the coordinate system of Fig. 2·2, there exists an electric field only in the y direction and magnetic fields only in the x and z directions. We represent these by

$$\begin{aligned} E_y &= A_y C e^{j(\omega t - kz)} \\ H_x &= G_x D e^{j(\omega t - kz)} \\ H_z &= G_z F e^{j(\omega t - kz)} \end{aligned} \tag{2·3}$$

where C, D, and F depend on x and will have to be determined by the boundary conditions. The remainder of the expressions represent wave propagation down the z dimension of the guide. Now if we apply equations 1·13a and 1·13b we find

$$\begin{aligned} E_y &= \frac{\omega}{kc} H_x \\ \frac{\partial E_y}{\partial x} &= -\frac{j\omega}{c} H_z \\ \frac{\partial H_x}{\partial z} - \frac{\partial H_z}{\partial x} &= \frac{j\omega}{c} E_y \end{aligned} \tag{2·4}$$

By some rearrangement, making use of the form of the expressions in 2·3 we obtain for one equation, involving E_y, the relation

$$\frac{\partial^2 E_y}{\partial x^2} = E_y\left(k^2 - \frac{\omega^2}{c^2}\right) \tag{2·5}$$

[3] McGraw-Hill Book Co., 1942.

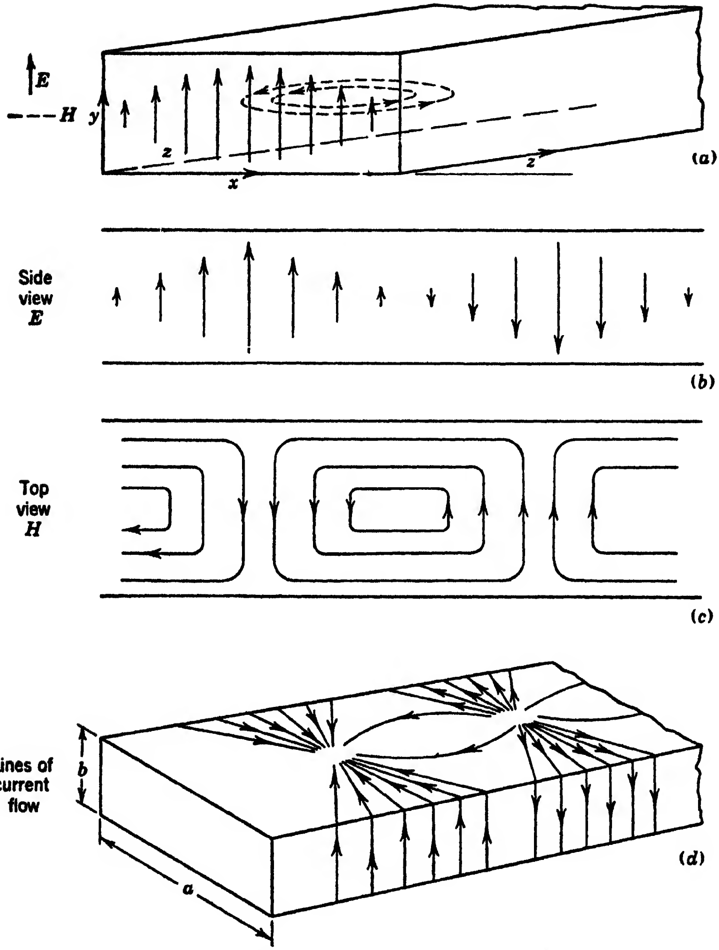


FIG. 2·2 Fields and currents in waveguides. The mode shown is the TE_{10} mode. The electric field is parallel to the short dimension, and the magnetic field runs around it. Currents flow up the sides into the middle of the wide dimension. The current is completed as displacement current due to the rapid time variation of the electric field across the guide.

This is a simple differential equation, which of course is satisfied by the expression for E_y in 2·3. The boundary condition for E_y is that it be zero at $x = 0$ and $x = a$. We can achieve this by making C take the form $\sin(\pi lx/a)$, where l is an integer. The second boundary condition applies the requirement that H_x is zero at $x = 0$ and $x = a$. This is again satisfied by making D take the form $\sin(\pi lx/a)$. Now the first two equations of 2·4 can be combined to give

$$\frac{\partial H_x}{\partial x} = -jkH_z$$

and this requires that F take the form $\cos(\pi lx/a)$.

We therefore find for the three fields

$$\begin{aligned} E_y &= A_y \sin \frac{\pi lx}{a}\, e^{j(\omega t - kx)} \\ H_x &= G_x \sin \frac{\pi lx}{a}\, e^{j(\omega t - kx)} \\ H_z &= G_z \cos \frac{\pi lx}{a}\, e^{j(\omega t - kx)} \end{aligned} \tag{2·6}$$

The fields and currents appropriate to the case where $l = 1$, designated as a transverse electric or the TE_{10} mode, are shown in Fig. 2·2. The electric and magnetic fields are shown together in (*a*), and separately in side and top view in (*b*) and (*c*). The lines of current flow are shown in (*d*). It is of interest to see that a slot along the center of the wide side of the guide does not interfere seriously with the current flow. On the other hand a slot down the length of the narrow side cuts across the current and produces a maximum disturbance of the wave. Thus it is possible to tolerate a slot in the wide side of the guide without much loss, but a slot in the narrow side results in considerable loss and indeed can be made the basis of an antenna feed, the so-called leaky waveguide feed.

Equations 2·5 and 2·6 are full of meaning. Consider first the quantity $k^2 - (\omega^2/c^2)$. This appears on the right in equation 2·5, and on substituting the expression for E_y from equation 2·6

we see that the width of the guide affects its value. For the simplest case, where $l = 1$, we get

$$k^2 = \frac{\omega^2}{c^2} - \frac{\pi^2}{a^2} \tag{2·7}$$

This relation shows a very important property of waveguides. The quantity k must be real, since it occurs in the exponential in a factor multiplied by j. If k is imaginary the two factors j give a *real negative term* in the exponential. This is interpreted as an exponential decay of the field strength. Substitution of values for an imaginary k in the wave equation show that the decay is extremely rapid.[4] Indeed it is similar to a process of reflection, in which the field is effectively rejected. Returning to our equation we see that this means that when π^2/a^2 is greater than ω^2/c^2 the value of k is imaginary and the waveguide *will not accept the electric and magnetic field at that frequency.* For a given value of a, the guide width, there is, therefore, a *cutoff frequency* given by $\omega^2 = \pi^2c^2/a^2$ or $f = c/a$. In terms of wavelength this appears compactly as

$$\lambda \leqq 2a$$

The free space wavelength must therefore be less than twice the guide width in this mode. We can apply the same reasoning to other modes. A most informative way of looking at this has been suggested by Hansen.[5] He points out that the second derivative with respect to a coordinate is the primary factor in the determination of the curvature of the field along that direction. Then the general requirement of an equation of the type

$$\frac{\partial^2 E_x}{\partial x^2} + \frac{\partial^2 E_x}{\partial y^2} + \frac{\partial^2 E_x}{\partial z^2} = \frac{1}{c^2}\frac{\partial^2 E_x}{\partial t^2}$$

which is the three-dimensional wave equation, is one of restricting the *total curvature* because $\partial^2 E_x/\partial t^2$ is determined by the frequency. Now if we proceed to fit a wave very tightly into a guide, the electric field has to be zero at the sides separated by the width a, and

[4] Suppose we wish to transmit 10,000 mc down a waveguide 1 cm wide. Then we have $k^2 = -(5/6)\pi^2$ or $k = 2.3j$. The field therefore drops to 1/2.7 of its value in 0.4 cm.

[5] W. W. Hansen, "Notes on Microwaves," Radiation Laboratory Report (1942).

this automatically supplies a curvature to the wave which gets greater the smaller we make the width. This high curvature must be compensated for in some way, and even if we put no additional restriction on the curvature we can see that the space dependence of the wave down the pipe, which is controlled by k, must adapt itself, since we suppose the frequency to be fixed. This is exactly what takes place. The wavelength increases as the guide is narrowed until at cutoff the wavelength is infinite.

The wavelength is given by $k = 2\pi/\lambda$. It is, for this simple mode, as deduced from equation 2·7, using λ_0 for free space wavelength and λ_c for cutoff wavelength,

$$\lambda = \frac{2}{\left(\dfrac{4f^2}{c^2} - \dfrac{1}{a^2}\right)^{1/2}} = \frac{\lambda_0}{\left(1 - \dfrac{\lambda_0{}^2}{4a^2}\right)^{1/2}} = \frac{\lambda_0}{\left(1 - \dfrac{\lambda_0{}^2}{\lambda_c{}^2}\right)^{1/2}} \tag{2·8}$$

Along with this change of wavelength goes a change in *phase velocity* and *group velocity*. The phase velocity is the velocity of arrival of the electric field. It is simply the product of frequency and guide wavelength, and therefore it is

$$v_{\text{phase}} = \frac{c}{\left(1 - \dfrac{\lambda_0{}^2}{\lambda_c{}^2}\right)^{1/2}} \tag{2·9}$$

The group velocity, the velocity of energy propagation, is given by $d\omega/dk$ and is, from equation 2·7, seen to be

$$v_{\text{group}} = \frac{c^2 k}{\omega} = \frac{c^2}{v_{\text{phase}}} = c\left(1 - \frac{\lambda_0{}^2}{\lambda_c{}^2}\right)^{1/2} \tag{2·10}$$

Notice that a signal travels with the group velocity which is lower than the velocity of light. On the other hand, the field arrives at points down the guide with the phase velocity and is faster than light. The last fact makes the design of some kinds of waveguide electron accelerators difficult because the material electron always falls behind the non-material field.[6]

[6] Note that the phase velocity is not a velocity of propagation because it is not measured along the direction of energy flow. The velocity of the point of intersection of a spherical wave with a plane can be infinite and represents the velocity of arrival of the field; not, however, measured along the direction of propagation of the wave.

If equation 2·6 is examined carefully it is seen that the fields can be duplicated by a fresh method of description. This has been developed by Adams and Page [7] and is a powerful method of describing the events in a waveguide. In place of a single field representation the field is ascribed to two plane waves traveling across the waveguide so that they are continually crossing and recrossing one another. It can be seen that this enables the representation of the curvature across the guide to be given. If the waves do not cross at all they are simply plane waves and there is no curvature. Of course, neither is there any guide. If

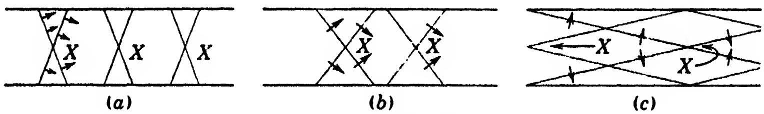


FIG. 2·3 Wavefronts in a waveguide: (*a*) shows the situation where the wavelength is short and the waves advance nearly down the guide; (*c*) shows the case where the wavelength is nearly double the guide width and the waves proceed nearly perpendicular to the direction of the guide.

they cross "head on" there is developed the maximum curvature of a standing wave. This is all the curvature that can be obtained; if it is not sufficient there can be no waves at all. The critical wavelength is double the guide width because the distance from node to node in a standing wave is half the wavelength.

Intermediate values of curvature can be arranged by letting the waves cross at lessening angles. This is shown schematically in Fig. 2·3. In (*a*) are shown the conditions when the width is ample. The waves travel nearly directly down the guide, and the phase velocity, which is the velocity which the points X develop, is not much higher than that of light in a vacuum. In (*c*) are shown the conditions near cutoff. The phase velocity is very great. The group velocity, which corresponds to the arrival of the plane waves themselves (having traveled many crisscross paths) is very small.

Description of other modes will be postponed to Chapter 4. We can add here a little about attenuation in waveguides.

[7] N. I. Adams and L. Page, *Phys. Rev.*, **52**, 647 (1937).

The fact that energy is dissipated in the currents which flow in the metallic walls determines the attenuation in waveguides. It can be supposed that the energy present in the skin depth represents the energy lost, and the rate of such loss per second can be thought of as the energy contained in a volume equal to the skin depth times the guide perimeter times the distance of travel in one second, v_g. To obtain an approximate figure it can be supposed that the entire energy in the skin depth is magnetic, which has been shown to be nearly true, and is due to a magnetic field which is half that at the surface. The energy lost per second is then

$$\left(\frac{c}{2\pi}\sqrt{\frac{\gamma}{\mu f}}\right) v_g \int_{\text{perimeter}} H^2\, dl$$

The energy transferred per second is $\frac{c}{4\pi}\int_{\text{area}} \mathbf{E} \times \mathbf{H} \cdot d\mathbf{S}$, the integral of the Poynting vector. Hence if dP/P is written for the ratio of power lost to power transmitted

$$\frac{dP}{P} = \frac{2\sqrt{\frac{\gamma}{\mu f}}\, v_g \int_{\text{perimeter}} H^2\, dl}{\int \mathbf{E} \times \mathbf{H} \cdot d\mathbf{S}} \tag{2·11}$$

This leads to an exponential absorption with the absorption coefficient derived from the two integrals of the above formula. Evaluating these for a rectangular waveguide leads to the expression for attenuation in decibels per foot:

$$\text{attenuation} = \frac{0.11}{a^{3/2}} \left[\frac{\frac{1}{2}\frac{a}{b}\left(\frac{f}{f_c}\right)^{3/2} + \left(\frac{f_c}{f}\right)^{1/2}}{\sqrt{\left(\frac{f}{f_c}\right)^2 - 1}} \right] \tag{2·12}$$

where f_c is the cutoff frequency, f is the frequency, and a and b are the wide and narrow guide widths in inches. Applying this equation to the standard guide used for 10-centimeter microwaves we have $f = 3000$ megacycles, $f_c = 1500$ megacycles,

roughly, and the result obtained is 5.6×10^{-3} decibel per foot. Accordingly one can transmit for 200 feet in such waveguide with a loss of only 1 decibel. This very low loss is due to the large ratio of volume to surface in the guide. Waveguide is clumsy but has its compensations.

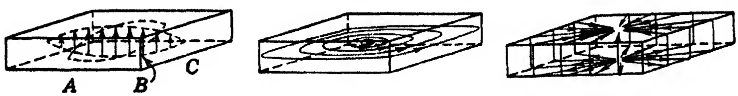
It is interesting that dirt does not greatly affect the loss in the guide unless it is partly conducting. A thin sheet of water in a guide gives very rapid attenuation.

2·3 RESONANT CAVITIES

The most firm restriction on the shape of an electromagnetic field is given by a closed conducting cavity. It is not at all surprising that only certain discrete frequencies can be maintained efficiently in such cavities. Of interest in addition to this fact is the size of field that can be developed and also its precise configuration, because both of these prove to be important in developing oscillators or building accelerators for electrons and protons.

First consider the effect of confining the field in a *rectangular cavity*. This shape is chosen solely for the purpose of illustrating how cavities behave. Rectangular cavities are rarely used in practice.

It has been seen in a waveguide that the boundary condition, that there be no tangential electric field at the conductor, is met by two waves traveling at an angle to the guide and crisscrossing continually. The only frequency requirement set up is that of a



(a) (b) (c)

FIG. 2·4 Showing (a) the electric field, (b) the magnetic field, and (c) the current in a simple mode of oscillation in a rectangular cavity.

minimum frequency which can be transmitted. Much the same process has to occur in a cavity, but there is now the added requirement that reflections occur from the ends and also from the top and bottom and yet a stable field pattern result. This can be followed in a simple case, and the general result can then be accepted. In Fig. 2·4 is shown a rectangular cavity carrying a

very simple mode of oscillation. This is achieved, as for a waveguide, by the reflection of two crossing waves first from the widely separated sides, second from the closely separated sides, the top and bottom being no concern in this mode. For a stable field pattern the distance C must be one-half guide wavelength for guide of width A, and, vice versa, A must be one-half guide wavelength for guide of width C.

Taking the values from equation 2·8 it can be seen that

$$\frac{\lambda_0}{\left(1 - \dfrac{\lambda_0^2}{4A^2}\right)^{1/2}} = 2C \quad \text{or} \quad \frac{\lambda_0}{\left(1 - \dfrac{\lambda_0^2}{4C^2}\right)^{1/2}} = 2A$$

Either of these leads to

$$\frac{1}{(2A)^2} + \frac{1}{(2C)^2} = \frac{1}{\lambda_0^2}$$

The general result which applies when reflections from all three pairs of faces are possible is

$$\left(\frac{1}{\lambda_0}\right)^2 = \left(\frac{f}{c}\right)^2 = \left(\frac{l}{2A}\right)^2 + \left(\frac{m}{2B}\right)^2 + \left(\frac{n}{2C}\right)^2 \qquad (2\cdot 13)$$

l, m, n are integers, and only one of the three terms on the right can be zero. The case where $l = m = 1$ and $n = 0$ is shown in Figs. 2·4 (*b*) and 2·4 (*c*). The electric vector is distributed as shown, the magnetic field is around the walls of the cavity, and the two field distributions are made possible by current flow as indicated in Fig. 2·4 (*c*). The current actually pulses back and forth. Only one surge is indicated. It can be seen that, if we imagine the top and bottom surfaces to be the plates of a condenser connected by a kind of distributed wire to carry the current flow, we have an analogy to an LC circuit in which the inductance is distributed over a number of wires.

The rectangular cavity we have described is not the only form. Many others exist. Some of these are shown in Fig. 2·5. The cylindrical form in (*a*) needs little further comment. Inspection of Fig. 2·4 will already have shown how the electric field tends to concentrate in the center of the cavity for the rectangular case. This is also true in the cylindrical case. The magnetic field is

around the circular part of the cavity. For such a mode of oscillation the resonant wavelength is given by

$$\lambda = \frac{4}{\sqrt{\left(\frac{l}{b}\right)^2 + \left(\frac{2.44}{r}\right)^2}} \tag{2·14}$$

where l is an integer, and r and b are as shown in Fig. 2·5 (*a*). This resonant wavelength is for this one type of mode only. For

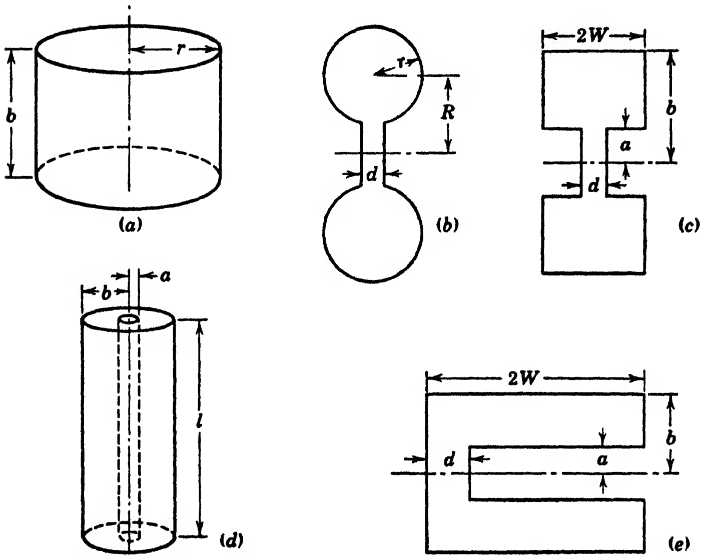


FIG. 2·5 Schematic diagrams of various types of resonant cavity.

other modes the second term in the denominator is changed. It is actually fixed by the solution of a Bessel equation.

The two "doughnut" oscillators shown in (*b*) and (*c*) are important because they approximate a very common form of cavity used in practice. In order to excite cavities directly by electron streams it is often necessary to keep small the distance traversed by the electrons. By building the cavity of the shape shown in (*b*) or (*c*) the electric field is concentrated in a small distance. This is the type of cavity found in a klystron (Chapter 3). The

resonant frequencies f in megacycles for these are given by

$$\text{Type }(b)\text{:}\quad \frac{3 \times 10^4}{f} = 2\pi\left\{\frac{\pi r^2 R}{d}\left(1 - \sqrt{1 - \left(\frac{r}{R}\right)^2}\right)\right\}^{\frac{1}{2}} \tag{2·15}$$

$$\text{Type }(c)\text{:}\quad \frac{3 \times 10^4}{f} = 2\pi\left(\frac{Wa^2}{d}\ln\frac{b}{a}\right)^{\frac{1}{2}} \tag{2·16}$$

where the lengths are measured in centimeters and are as shown in Fig. 2·5.

The coaxial resonator shown in (*d*) is of interest since it provides a very strict limitation on the mode of oscillation. The field is limited to the form of radial electric and circular magnetic fields. Waves of frequency such that they will fit as standing waves into this pattern, having E zero at each end of the cavity if it is closed off, can excite such a cavity. The first mode of oscillation is accordingly that for which the wavelength is twice the length of the cavity. This is so simple a mode of oscillation that such cavities are chosen as wavemeters. Note here the analogy with acoustic resonators which are tubular in shape.

Substitution of numbers in any of these expressions shows that cavities tuned to frequencies up to 30,000 megacycles ($\lambda = 1$ centimeter) are readily obtainable.

Energy Storage in Cavities

We have already seen that the boundary conditions for confining electric and magnetic fields are not perfect. The penetration of the field into a metal is not great but it is finite. This fact has an important bearing on the behavior of cavities. Consider a perfectly conducting cavity. Imagine energy fed in slightly off resonance. In order for this energy to remain in the cavity a form of oscillation which is not perfect must exist; there must be a slight value of E at a surface where ideally it should vanish. If the conduction is perfect this slight value of E invokes huge currents which continue as long as E attempts to maintain its value. These huge currents can be looked on as rejecting power, with the result that the cavity refuses to accumulate energy. On the other hand, if the cavity walls have some resistance there will be a minimum value of E below which there will be inadequate current to cause the vanishing of E, because a certain current requires a certain

field. Up to this not very definite value the cavity will be able to accept energy. This means that the cavity becomes less selective of frequency, which may or may not be desirable depending on what the cavity is to be used for.

Selectivity is a commonplace in radio engineering. It has become usual to describe the selectivity of an oscillatory circuit in terms of a quantity Q which is variously defined. Sample definitions are: $f/2Q$ is the number of cycles off resonance for 70.7 per cent of peak current (this is a definition directly in terms of the resonance curve); $Q = \omega L/R$, where L is the inductance, R the resistance in an oscillatory circuit, and ω the angular velocity of vector rotation describing the frequency of oscillation. This definition is useful for direct substitution in equations of circuits. The first definition is, of course, applicable, but it does not help in making estimates of the behavior of cavities; the second is hard to use. There is an equivalent definition: *Q is 2π times the ratio of the energy stored in the circuit to the energy lost per cycle.* It is equivalent to stating that the rate of loss of energy is given by

$$\frac{\text{energy at time } t}{\text{initial energy}} = e^{-\omega t/Q}$$

This is readily applicable to cavities.

The energy stored in the free space of the cavity is $(1/8\pi)(E^2 + H^2)dV$, where V denotes volume. We have pointed out that some energy penetrates into the walls of the cavity in the region of the skin effect, and we can suppose that this represents the energy which is being bled off. As we start oscillations in the cavity the energy we supply goes to increasing the energy in the space and only in small part to the walls. However, as the energy stored increases so does the loss due to the skin effect until a limiting condition is reached where the energy provided in a new cycle of feeding in power is all lost in the walls. This quantity can either be calculated by a fair amount of mathematics or measured experimentally. It is possible to get some idea of Q for a cavity, and also of the factors that operate to improve or lessen it, by assuming that the average value of H is one half that at the surface and that the energy in the volume of the walls defined by the skin depth is lost. We also assume that E and H share energy equally. Then we get for the energy stored $(1/8\pi)(2H)^2V$, where V is the cavity volume, and for the energy lost $(1/8\pi)(2H)^2S_T\alpha$, where

S_T is the surface area of the cavity and α is the skin depth. We thus find the approximate relation

$$Q = \frac{\pi V}{S_T \alpha} \tag{2·17}$$

A cylindrical cavity with a volume of 100 cubic centimeters has a wall area of about 200 square centimeters. At 10 centimeters wavelength the skin depth is about 10^{-4} centimeter, so that a Q of $100\pi/(200 \times 10^{-4})$, or 15,000 roughly, would be expected. This is much higher than the values of Q attainable in the kilocycle region by the use of coil and condenser tuned circuits. For such circuits a Q of 500 is good. Indeed, one of the problems of microwave practice is that of avoiding the high selectivity which is provided by large metal surfaces.

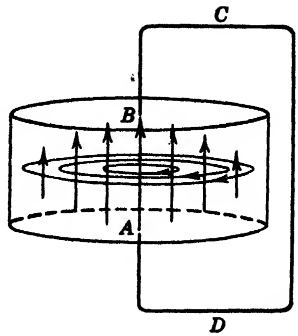


FIG. 2·6 The voltage developed in a cavity. The circuit is in a region of zero field except within the cavity where it runs along the highest electric field. Since it encloses the whole magnetic flux the time rate of change of the magnetic flux enables the voltage to be found.

Voltage Developed in Cavities: Shunt Resistance

For many purposes a cavity is required to develop a voltage which is used for a secondary purpose. For example, the cavity may be part of a microwave oscillator and so may have to control an electron stream. Or it may be part of a nuclear accelerator with the aim of developing high energy protons or electrons. It is therefore important to have some figure associated with a particular kind of cavity which will enable the voltage developed to be calculated.

The voltage developed is calculable from equation 1·6. This, in words, states that the emf around a circuit is equal to the rate of change of magnetic flux threading the circuit. The circuit we can choose is shown in Fig. 2·6, where we take the whole of the electric field in going across the cavity and enclose the whole magnetic flux also. Since we are dealing with air which has $\mu = 1$ we have

$$\text{Voltage} = \int E\,dl = -\frac{1}{c}\int \frac{\partial H}{\partial t}\,dS \text{ statvolts} \tag{2·18}$$

This enables us to calculate the voltage for any given value of the electric and magnetic fields in the cavity. Actually this is not of very great use because we are really interested in knowing the voltage which we can develop when the power we feed in is in equilibrium with the power lost in the walls. The considerations of the previous paragraph showed that the energy lost per cycle is $(1/2\pi)H^2S_T\alpha$. If the frequency is f the energy lost per second is then $(f/2\pi)H^2S_T\alpha$. Now we can define the shunt resistance R_s by using the familiar relation

$$\text{power lost} = \frac{(\text{voltage})^2}{R_s}$$

in which case we find that

$$R_s = \frac{(\text{voltage})^2}{\text{power lost}}$$

$$= \frac{\dfrac{2\pi}{c^2}\left(\displaystyle\int \frac{\partial H}{\partial t}\,dS\right)^2}{H^2S_T\alpha f} \tag{2·19}$$

For rough thinking we can see that, if we deal with averages, the average of $(\partial H/\partial t)^2$ is roughly $4\pi^2f^2H^2$ since H is of the form $Ae^{2\pi j[ft-(x/\lambda)]}$. The area of a slice across the cavity and the area of the periphery are geometrically related; therefore we can reduce R_s to a quantity of the form

$$R_s = \text{constant} \times \text{area} \times \frac{f}{\alpha} \tag{2·20}$$

where the area term includes the geometrical relation mentioned above. The relation is, of course, different for different shapes of cavity.

Values for R_s are given in Table 2·1. They are around a half megohm. This high figure shows that a microwave cavity is an important source of voltage. If as little as 1 watt is fed into a cavity with a shunt reistance of 1 megohm it will develop 1000 volts at the position of maximum voltage. The control of an electron stream by this voltage is therefore no slight matter; in fact

it exceeds the control voltages used in all but the highest power radio tubes. This point is made here because one's first impression of microwave technique is that it involves small voltages and low powers. The reverse is true. The high voltages developed in cavities render possible the control of large amounts of electronic energy, with the result that very large amounts of r-f power can be developed.

Cavities in Practice

A common use of cavities in practice is the development of voltage, but this is not the only use. Cavities are used as tuning devices, as wavemeters, and as filters. One of the most frequently used cavities is the elongated doughnut, or loaded cylindrical cavity, shown in Fig. 2·5 (*e*). This shape develops the voltage variation across a relatively narrow gap, which permits the control of an electron beam without too much trouble from the finite velocity of the electrons. In addition it is easy to construct. The actual resonant frequencies for various values of the radii and lengths have been worked out. Some rather heroic approximations are given in Table 2·1. Fuller information is given in several sources.[8]

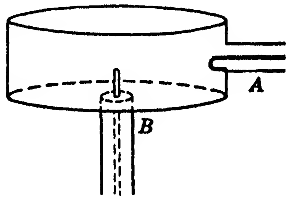


Fig. 2·7 Excitation of a cavity by a loop or a probe. The loop produces a magnetic field where the oscillation produces the highest magnetic field. The probe excites the vertical electric field.

The excitation of cavities is important. It is necessary that the type of excitation have at least a component of either the electric or the magnetic field in the direction required by the cavity oscillation. Thus to excite a cavity such as a cylinder, a loop like that indicated at *A* in Fig. 2·7 may be employed. A loop at position *B* is not satisfactory because the magnetic field of the cavity oscillation does not thread the loop. For the same reason a probe at position *A* does not work. Excitation by waveguide is generally at position *A* with

[8] See for example *Microwave Transmission Design Data*, Sperry Gyroscope Company (1944), pp. 198 et seq., or Terman, *Radio Engineers Handbook*, McGraw-Hill Book Co., 1943, pp. 264 et seq.

TABLE 2·1 FEATURES OF CAVITIES

(α is the skin depth in the appropriate units.)

Type of Cavity	Resonant Wavelength (λ)	Q	R_s
Rectangular box as in Fig. 2·4	$\dfrac{2}{\sqrt{\left(\dfrac{l}{A}\right)^2 + \left(\dfrac{m}{B}\right)^2 + \left(\dfrac{n}{C}\right)^2}}$	When $A = C$, $n = 0$, $1 = m = 1$ $0.553 \dfrac{\lambda}{\alpha} \left(\dfrac{1}{1 + \dfrac{A}{2B}}\right)$ When $A = B = C$ and mode is high, $\dfrac{A}{4\alpha}$	As for Q $340 \dfrac{B}{A} Q$
Cylindrical resonator as in Fig. 2·5 (*a*) TE modes	$\dfrac{4}{\sqrt{\left(\dfrac{l}{b}\right)^2 + \left(\dfrac{2u'_{n,m}}{\pi r}\right)^2}}$ $u'_{n,m}$ is the mth extreme of $J_n(u)$, or the mth root of $J'_n(u) = 0$ $u'_{0,1} = 3.832$, $u'_{1,1} = 1.841$ $u'_{0,2} = 7.016$, $u'_{1,2} = 5.330$		

TM modes	$\dfrac{4}{\sqrt{\left(\frac{l}{b}\right)^2 + \left(\frac{2u_{n,m}}{\pi r}\right)^2}}$ $u_{n,m}$ is the mth root of $J_n(u) = 0$ $u_{0,1} = 2.405,\ u_{1,1} = 3.832,\ u_{2,1} = 5.136$ $u_{0,2} = 5.520,\ u_{1,2} = 7.016,\ u_{2,2} = 8.417$ $u_{0,3} = 8.654,\ u_{1,3} = 10.173,\ u_{2,3} = 11.620$	$n \neq 0$ $\dfrac{r}{\alpha}\dfrac{1}{1 + \frac{r}{b}}$ $n = 0$ $\dfrac{r}{\alpha}\dfrac{1}{1 + \frac{r}{2b}}$	For $l = n = 0,\ m = 1$ $R_s = \dfrac{376b}{\alpha} \cdot \dfrac{1}{1 + \frac{r}{2b}}$
Coaxial line resonator as in Fig. 2·5(d)	$2l$	$\dfrac{\lambda}{\alpha}\dfrac{1}{4 + \frac{l}{b}\left[\frac{1 + (b/a)}{\ln(b/a)}\right]}$	$\dfrac{60\lambda b[\ln(b/a)]^2}{\pi\alpha l\left(1 + \frac{b}{a}\right)\left\{1 + \frac{b[\ln(b/a)]}{l[1 + (b/a)]}\right\}}$
Loaded cylinder as in Fig. 2·5(e). All values are approximate only	$2\pi\sqrt{\dfrac{Wa^2}{d}\ln\dfrac{b}{a}}$ For $\dfrac{2W}{d}$ large.	For $a = 0.016\lambda$, $\dfrac{d}{a} = 0.5,\ 1.12 \times 10^3\sqrt{\lambda}$ $1.0,\ 1.34 \times 10^3\sqrt{\lambda}$ $2.0,\ 1.51 \times 10^3\sqrt{\lambda}$ For $a = 0.063\lambda$, $\dfrac{d}{a} = 0.5,\ 1.56 \times 10^3\sqrt{\lambda}$ $1.0,\ 2.34 \times 10^3\sqrt{\lambda}$	 $1.40 \times 10^5\sqrt{\lambda}$ $1.96 \times 10^5\sqrt{\lambda}$ $2.51 \times 10^5\sqrt{\lambda}$ $1.12 \times 10^5\sqrt{\lambda}$ $2.24 \times 10^5\sqrt{\lambda}$
Spherical of radius r	$2.28r$, First mode $1.4r$, Second mode	$0.318\dfrac{\lambda}{\alpha}$, First mode	$104.4\dfrac{\lambda}{\alpha}$

the guide turned so that the electric vector in the guide is parallel to the electric vector in the cavity.

Designation of Modes

The various modes of oscillation in a cavity are designated by numbers. A mode is referred to an axis of some kind. The field which is not in this axis, and which is accordingly transverse to it, is then used to designate the mode. Thus, if we pick the axis of a short cylinder as the reference axis and if the electric field is along this axis, the magnetic field is necessarily transverse to it. Such a mode is called a transverse magnetic or TM mode. A mode in which the electric field is along radial lines has no electric field along the axis and so is called a transverse electric or TE mode. The numbers which pertain to the appropriate solution of the field equations are added as subscripts.

These numbers are exactly analogous to quantum numbers in atomic theory. They are not so important because it is generally vital to eliminate all but the lowest mode to produce the desired operation. Consequently it is not necessary to memorize a complicated set of designations. One point is worth making, however, and that is that the numbers used in mode designations for different geometrical shapes do not designate the same field configuration. Thus a mode which is TE_{10} in rectangular waveguide maintains very nearly the same configuration if a taper join to a cylindrical guide is made. It is there designated as TE_{11}.

Summary

The nature of electrical oscillations in coaxial line, waveguide, and cavities has been considered. The boundary conditions greatly limit the form taken by the fields. In coaxial line it requires a special field distribution. In waveguide it necessitates a certain minimum wavelength and also certain definite field configurations. In cavities the limitation takes the form of suppressing all but certain definite frequencies.

PROBLEMS

2·1 Find the change in resonant frequency of a rectangular copper cavity which is cooled from 20°C to −258°C, assuming the skin depth to be added to the dimensions. Take $A = B = 5$ cm, $C = 1$ cm, and treat the lowest mode.

2·2 Estimate the cost of gold-plating 100 ft of 3 by 1½ in. waveguide to the skin depth at 10 cm free space wavelength.

2·3 Calculate the guide wavelength and phase velocity in a waveguide 3 in. wide at 4000, 3000, 2500 mc.

2·4 Calculate the peak voltage developed in a cylindrical cavity oscillating in a TM mode, $l = n = 0$, $m = 1$ when 1 megawatt of power is fed in. Take $r = 20$ cm, $b = 10$ cm.

CHAPTER 3

THE PRODUCTION OF MICROWAVES

There has for a long time been a considerable interest in the production of very high frequency radiation. It was realized almost without trial that the conventional resonant circuit so widely used in the production of radio broadcast frequencies would not be suitable for the production of microwave oscillations. With a conventional circuit having a capacitance of 1 micromicrofarad and an inductance of 1/100 microhenry, both of which are obtained only by the utmost care applied to the reduction of stray effects, the resonant frequency is about 2000 megacycles per second, which is by no means high in the microwave region.

Some success was obtained with parallel wire lines attached to triodes in which the electrons performed oscillations about the grid, the so-called Barkhausen oscillations. The excitation of iron filings suspended in oil was also studied as a means of exciting short wave radiation. The remarkable work of Cleeton and Williams in 1934 was performed with split-anode magnetrons and very short, enclosed, Lecher wire oscillators. However, no great success was attained until attention was turned to the excitation of oscillations in cavities.

The first published account of reasonably satisfactory microwave oscillators came in 1938 from Stanford University, where analyses of cavity oscillations were made by Hansen, and the velocity modulation tube, known as the klystron, was designed by the Varian brothers. The idea of exciting cavities was taken up

in several places and there resulted the "lighthouse tube" triode, developed by McArthur at the General Electric Company, which can produce very effective oscillations in the longer microwave region; and the cavity magnetron which amazed the Anglo-American scientists at the onset of the war by its ability to deliver enormous power when put into pulsed operation. The magnetron development came from Birmingham University in England, and the credit goes to Oliphant and his group including Randall, Boot, and Duke.

This chapter comprises an account of these three methods of exciting cavity oscillations and a description of the major characteristics of the tubes which are used for the purpose.

3·1 MAINTENANCE OF OSCILLATIONS IN CAVITIES

The most effective method of supplying energy to maintain oscillations is to control the motion of electrons. These are readily responsive to high frequency fields on account of their light mass, and at the same time they can produce a relatively large effect on account of their charge. The simplest manner of using electrons is shown in the familiar triode oscillator illustrated in Figure 3·1.

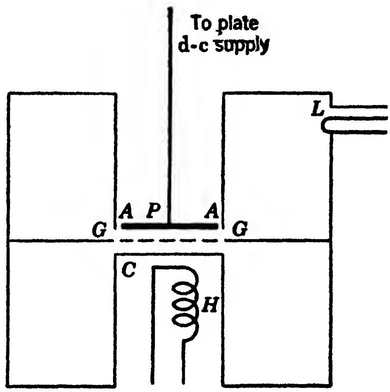


FIG. 3·1 Illustrating the mode of action of a simple cavity oscillator consisting of two resonant cavities coupled at the grid of a triode.

Two re-entrant cavities are placed back to back. The surface of one is coated with electron-emitting material and is heated by a hot coil to provide the electrons. The surface opposite is made into a grid. The space charge between the cathode C and the grid G is therefore, as in a normal triode, susceptible to the effects of an r-f voltage on the grid. The re-entrant surface of the second cavity is cut free from the rest of the metal and is insulated from it by a

very thin layer of insulator such as mica. This layer actually changes the structure of the cavity but it can be tolerated because the region near the gap of the cavity is one in which conduction current does not flow heavily, but there is a large and rapidly changing electric field. A reference to the currents in the rectangular cavity illustrated in Fig. 2·4 shows that this is so. This point may also be expressed by the statement that the requirement that the normal electric field end on the re-entrant plate is satisfied if there is free charge there. This charge can be supplied from the external circuit if necessary. The reason for this insulation is to permit the plate P to be maintained at the required d-c potential to operate as a component of a triode.

Since the grid is shared by both cavities there is considerable coupling between them and this coupling is favorable to oscillations. If a situation is imagined where the grid goes slightly positive, there will be a flow of electrons to the plate, which will cause the plate to go more negative. The field across the gap then increases and so builds up the flow of electrons until the cycle ceases because the supply of electrons runs out. In turn, a rise is caused in the plate voltage, and the whole process reverses until the electron flow is shut off by the negative voltage of the grid. The action is just like that of a tuned plate-tuned grid oscillator. The power is fed out from the upper cavity by a loop terminating a coaxial cable. The loop is threaded by the magnetic field in the cavity, and a voltage is excited in the conductor which starts r-f energy going down the cable.

Notice two design features which are common in r-f technique. The first is the isolation of d-c voltage with thin mica which does not interfere with the radiofrequency currents; the second is the pickup loop formed at the end of a coaxial cable which is the commonest method of extracting r-f power from an oscillating cavity.

It is obvious that there is nothing new or startling about this type of oscillator. Why then has it taken so much research to complete the development? The answer lies primarily in the necessity of maintaining the required *phase relationship* between the field and the electrons, and secondarily in the need for construction of a special tube to fit the two cavities and of methods of tuning. These can now be discussed, but the discussion will be clearer if we consider an actual tube and how it operates.

In Fig. 3·2 is a schematic drawing of a GL446 lighthouse tube in a coaxial cavity oscillator; E. D. McArthur is responsible for the design. This arrangement is electrically the same as the circuit in Fig. 3·1 but the grid cavity is folded back around the plate cavity to make possible the insertion of the tube and also to simplify the tuning adjustments. The tube is designed so that cath-

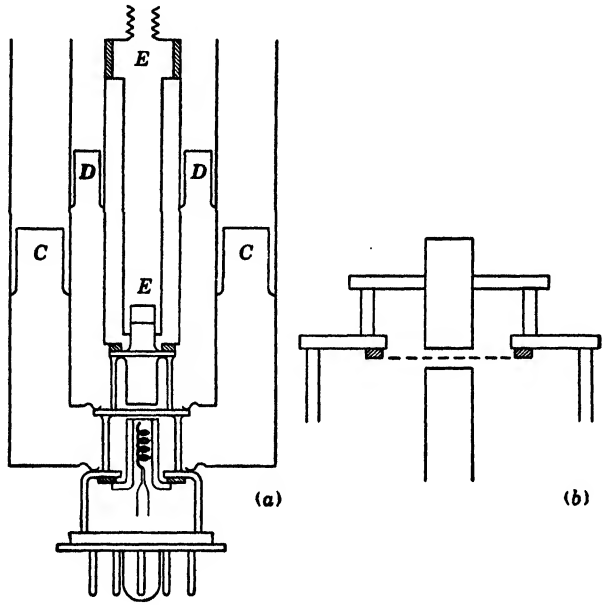


FIG. 3·2 Schematic drawing of a lighthouse tube and associated cavities.

ode, grid, and plate can be attached to spring fingers forming part of the metal cavities. The heater is supplied from a conventional octal base. Tuning is accomplished by moving the two shorting plungers *CC* and *DD*. The plate is insulated by means of a mica ring at the top of a long plate lead running down the inner conductor of the plate cavity. The cavities are generally set at about the required positions by measurement, and then fine tuning is achieved by a slight motion of the plate rod in and out, the plate connector on the tube having a little extra length to allow this.

We can now discuss some of the design problems. There are three major ones, namely phase relationships, power dissipation, and mechanical cavity construction.

Phase Relationships. Two factors operate to spoil the correct phase relationships between the electrons and the fields. The first is *interelectrode capacitance*, and the second is *transit time.*

At these frequencies the reactance of small capacities is considerable, for example 17 ohms for a 1-micromicrofarad capacitance at 10,000 megacycles. If such a capacitance shorts the grid and plate because of any factor not actually part of the oscillating cavity a phase shift will be introduced which can disturb the arrival of the electrons at the plate and reduce efficiency to a point where the tube will not operate. Hence tubes must be designed with small interelectrode distances so that stray capacitance effects are minimized. Up to oscillating frequencies of around 3000 megacycles this is the most important factor in design. Above 3000 megacycles the effect of transit time becomes of paramount importance.

Transit time is a rather general term given to the time taken by electrons to move between electrodes. It is usually wholly negligible, but not at microwave frequencies. This can be shown as follows: the relation giving the velocity of an electron is derived from $\frac{1}{2}mv^2 = Ve$, where m is the mass of the electron in grams, v the velocity in centimeters per second, V the potential drop through which the electron falls in statvolts, and e the electronic charge in statcoulombs. Reduced to a relation between volts and centimeters per second this becomes

$$\text{Velocity} = 6 \times 10^7(\text{volts})^{\frac{1}{2}} \tag{3·1}$$

This relation holds, of course, only for low velocities where the mass is effectively the rest mass.

Suppose that the grid and plate are separated by 1 millimeter and that there is a potential difference of 100 volts between them. If an electron moves across, the time of flight is $\frac{1}{3} \times 10^{-9}$ second. At 1500 megacycles the time of one whole oscillation is $\frac{2}{3} \times 10^{-9}$ second, or twice as long. This would be most unfortunate from the point of view of maintaining oscillations, for the plate would receive current at a phase exactly opposite to what it

called for at the time the electrons were detached from the space charge. The distance of 1 millimeter is therefore much too great for the grid plate spacing at 1500 megacycles. This is one new kind of limitation on microwave triode oscillators.

To describe the transit time in terms which show its relation to the frequency, it is generally expressed as a *transit angle*, which is simply the phase angle occurring in the time required by the transit of the electrons. In the case above it is π radians or 180°. The transit angle increases roughly linearly with frequency. Thus if a tube is constructed with spacings of the order of a third of a millimeter it might oscillate satisfactorily at 2000 megacycles but would have five times the transit angle at 10,000 megacycles and very probably would not oscillate.

Power Dissipation. The second consideration listed above is the power dissipation. A vacuum tube with small interelectrode spacings must have a fine wire grid and will therefore have a very poor design for dissipating heat. Accordingly it cannot be expected that under continuous oscillation these triodes will give out more than the order of a watt. Under pulsed conditions where the tube is on for only a fraction of the time the peak power output may be considerably more—of the order of a kilowatt. Actually the experimental utility of this kind of triode may be very great. For nuclear accelerators the frequency required is lower than was found optimum for radar development. At about 50 centimeters wavelength the effects of both transit time and stray capacitance are very much less, and sturdier construction is possible. Already tubes are available in experimental form which are potentially able to give a megawatt of pulsed power at low ratios of "on time" to "off time."

Cavity Construction. The third consideration, cavity design, naturally depends on the use to which the tube is to be put. As a laboratory oscillator the cavity design can require careful machining. For production in large quantities, as part of a mass-produced radar beacon, for example, this is not feasible, and much thought must be given to the design. The method of using coaxial cavities and sliding tuning as shown schematically in Fig. 3·2 seems to be successful. The tube can also be applied to more conventional cylindrical cavities.

3·2 VELOCITY MODULATION: THE KLYSTRON

The apparent disadvantage of transit time is made use of in velocity modulation oscillators, of which the klystron, invented by the Varian brothers,[1] is the best known. The klystron has been considerably developed by the Sperry Gyroscope Company and Bell Telephone Laboratories, with help from the Radiation Laboratory, and is now one of the most useful and adaptable tubes for microwave research. It operates on a totally different principle from the ordinary triode in that space charge is no factor in the operation and also in that the phase of the electrons is controlled by controlling their time of flight.

Stated briefly, the klystron lets the r-f field of a cavity act on a stream of electrons. The uniform stream is then broken up into regions of large and small density. This "bunched" electron beam is then reflected back through the same cavity, the time being controlled by the reflector voltage to cause still more oscillations in the cavity. The great advantage is the lack of necessity for small spacings; a second advantage is the ability to control the frequency of the oscillations by manipulating the phase of the electron beam.

A schematic drawing of a reflex klystron is shown in Fig. 3·3. A cavity of the re-entrant type is made of material suitable for sealing to glass. A gun structure is sealed to one end and the reflector plate to the other. The cavity is maintained at ground potential while the gun structure is held at suitable voltages to deliver a beam of electrons through the narrow part of the cavity. This is drilled out and a grid structure fastened to it to permit the flow of electrons. The electrons are then repelled back through the cavity by a negative voltage on the reflector plate. Since any variation in voltage on either the reflector or the gun structure will change the speed of the electrons and therefore their phase relationships, it is necessary to use a well-regulated power supply. The design of these is discussed in Chapter 10. The klystron is tuned with respect to coarse adjustment over, say, 50 megacycles by means of cavity distortion. The cavity is made of flexible metal and can be put out of shape by adjusting three tuning screws. After this has been done and the klystron has been

[1] R. H. Varian and S. F. Varian, *J. Applied Phys.*, **10**, 321 (1939).

brought to the right range of frequency there is usually enough control over the frequency by means of the reflector voltage to complete the tuning. Power is fed out by the usual method of a loop and coaxial cable. A waveguide can also be fed by an "iris" output of the type described in Chapter 4.

The general principle of operation of the klystron is simple; the detailed behavior is more complicated. The first point to understand is *bunching*. The electrons which traverse the gap in the

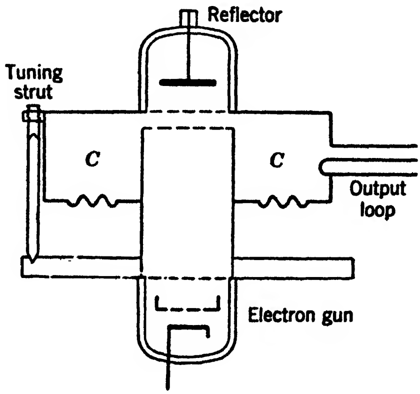


FIG. 3·3 Schematic drawing of a reflex klystron. The electron stream is modulated by the cavity oscillations on the way out, and on its return it is so bunched that the field of the electrons increases the amplitude of oscillation.

absence of oscillations will be roughly uniform in density [2] and will return in the same manner. There will be no sustained periodic variations of voltage in the gap due to irregularities in the electron stream. But if an r-f field exists, the electrons will be accelerated through the gap in one phase, while in the other they will be retarded. This will cause a relatively low density in between, for electrons will travel faster out of this region and will fall behind because of retardation. Where the faster electrons catch up to a previously retarded group there will be bunches of higher density. The formation of regions of high density is not too simple to de-

[2] The non-uniformity is due to the random emission of electrons from the filament. This emission is not modified as in a triode by accumulation to give space charge, but is present to its full extent. A klystron is therefore somewhat "noisy" and cannot be used as an amplifier where signal-to-noise ratio is important (cf. Chapter 8).

scribe numerically as it depends on the original velocity, the value of the r-f field, and the drift distance after the gap. However, when this bunching exists it is clear that the current flowing out through the gap is not as a rule the same as that flowing back, and so there will now be periodic variations of field in the gap due to the electrons as well as to the cavity oscillation. If these periodic variations can be made such as to maintain oscillations, the tube works. Considerable control can be exerted by the reflector voltage which determines the drift time and, therefore, the phase of the field due to the electrons.

A certain amount of understanding of the process of bunching can be obtained from Fig. 3·4. In (*a*) is shown a representation of a klystron with the gun structure represented by C, the grids at the cavity by G_1 and G_2, and the reflector by R. The d-c-voltage distribution through the tube is indicated in (*b*), with reversed sign, to allow for the negative sign of the electron. In (*c*) is indicated the variation of r-f voltage across the grids with time, the cycle being divided into eight equally spaced intervals. If the time to enter the reflector field, turn around, and arrive at any given point is AA_1, in the absence of r-f field the electrons which are in the gap at the eight equally spaced intervals will arrive equally spaced. However, if an electron at the time represented by 2 is accelerated in going across the gap, it falls farther into the repelling field and takes longer to return to the point considered, just as a stone thrown upward takes longer to return to the ground. This longer time is indicated by placing the dot farther to the right to signify more time. Similarly an electron crossing the gap at time 6 is retarded, enters the repelling field slower, and quickly returns. This dot is placed to the left of normal. The closer spacing of dots 3, 4, and 5 is clearly seen, and it corresponds to bunching. There is lower electron density for phases 7, 8, and 9.

The theory of bunching has been worked out by Webster,[3] whose papers should be consulted for further details.

The next point to consider in the klystron oscillator is the way in which the electrons maintain oscillation. From the point of view of energy conservation this is easy. The electrons must arrive back at the grids when the r-f field is opposing their motion, for

[3] D. L. Webster, *J. Applied Phys.*, **10**, 501 (1939); 864 (1939).

then they will lose energy, and presumably the energy goes to maintaining the electric field so that the oscillations continue or even grow. If the reader wants to *see* the process rather than

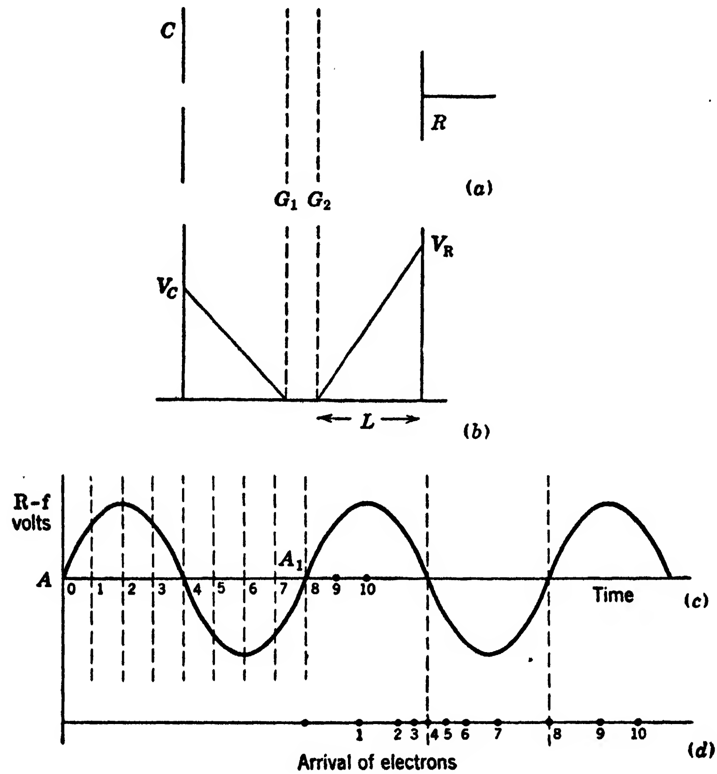


FIG. 3·4 Illustrating the process of bunching: (*a*) is a representation of the cathode and electron gun, the two grids at the cavity, and the reflector; (*b*) shows the potential field in which the electrons move; (*c*) shows how the time of arrival of electrons which have been reflected depends on the phase at which they start.

examine an energy balance sheet, then the following will help him.

If the two opposing streams of electrons are subtracted from one another there remains a "slug" of electrons which traverse the cavity. If these are retarded by the field they spend more time where they move slower, near the negative plate of the gap.

They themselves, however, exert a field while lingering in the neighborhood of the negative grid, and this increases the field in the gap, which in turn increases the flow of current in the walls of the cavity. An increased oscillation in the cavity results even after the electrons have gone through the grid. On the other hand, if the electrons are accelerated by moving toward a positive grid they are rapidly removed from the region where they are aiding the r-f field.

An interesting way of looking at a velocity modulation tube is due to Knipp, who points out that the passage of the bunch across the gap represents a current turned on for a short time. The application of the Fourier integral to this gives a frequency peak with subsidiary maxima. Thus for a pulse confined to 10^{-9} second the frequency peak is at 1000 megacycles. According to Knipp, if this peak is close to the oscillating frequency the electrons can drive the cavity; if the phase is right it will happen. If this idea is applied to the actual tube the general behavior can be predicted.

There are now some experimental features of klystrons which require discussion.

Klystron Cavities: Tuning Range

One very serious problem of microwave oscillators is that of varying the frequency over even a moderate range. To do this for a cavity oscillator necessitates changing its shape. One method makes the cavity of thin flexible walls and distorts these. The distortion changes the whole operation of the cavity, but if its construction is of the re-entrant type, where the modes of oscillation are largely determined by the already odd geometry, the mode does not change but the frequency does. This method can give about a 10 per cent frequency variation, though it is not adaptable to continuous tuning because the klystron goes out of oscillation unless the reflector and cathode voltages are varied as required to keep the tube going. A second method uses wide screws which are inserted in the cylindrical walls of the cavity. By screwing these in and out the cavity can be tuned, again over a range of about 10 per cent. This type of cavity is rather heavy and is advantageous only with klystrons like the 707A which are adapted to external cavities.

Electrical Tuning

The oscillations of the klystron are not entirely governed by the cavity, though it is paramount. The electron beam also takes a part in the oscillation, and if the amount and time of arrival of the bunches are varied the frequency of the oscillations can be controlled to some extent. This procedure is very useful in practice for it enables the frequency of the oscillator to be swept by an electronically controlled voltage which can be derived from an automatic frequency control discriminator or locked in with a sweep on an oscilloscope. In the first case, one can obtain automatic tuning (Chapter 8) and, in the second, frequency scanning as in a spectrum analyzer.

The degree to which the reflector affects the frequency of the oscillator depends on the internal construction and the frequency. In the 707A a reasonable power output is obtained over a range of about 25 megacycles. In the 723A, which operates in the 3-centimeter region, the frequency change per volt on the reflector is much greater. Several values of the reflector voltage will cause oscillation. This follows naturally from the fact that more than one manner of bunching will cause electrons to arrive at the gap in the cavity at the right time. In practice one value generally gives considerably more power than the other modes.

Velocity-Modulated Tubes in Practice

At 10 centimeters klystrons can give about 100 watts continuous output. This assumes that they are adequately cooled and that the voltages are high enough. A double-cavity klystron is more suitable for cooling as there is no need for a reflecting electrode. Also a "straight-through" technique is slightly more efficient in producing the optimum bunching. Such klystrons may be useful in the laboratory. They did not find much application in pulsed radar because they are harder to tune. In radar use as local oscillators in superheterodyne receivers klystrons give about 500 milliwatts and operate with less than 500 volts on any electrode. The upper limit of power is much less at 3 and at 1 centimeter, but there is still adequate power to give efficient frequency conversion in a crystal mixer in a superheterodyne receiver.

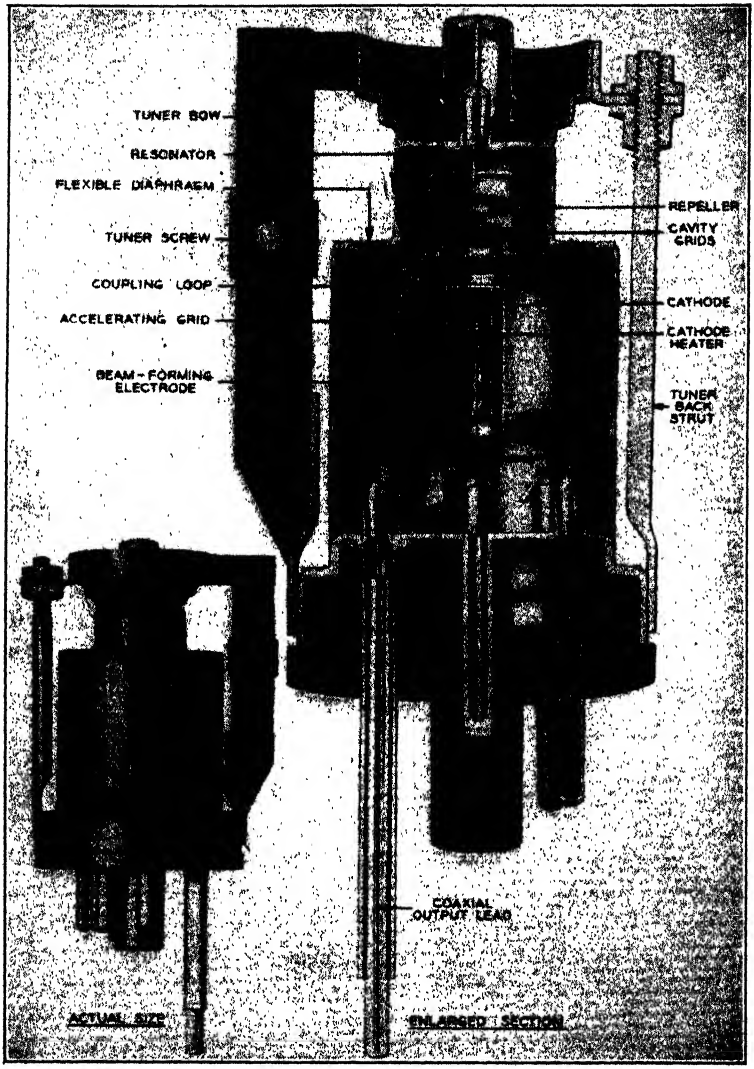


FIG. 3·5 Cut-away view of the 723A tube, which operates at 3 cm wavelength. (Courtesy of Bell Telephone Laboratories.)

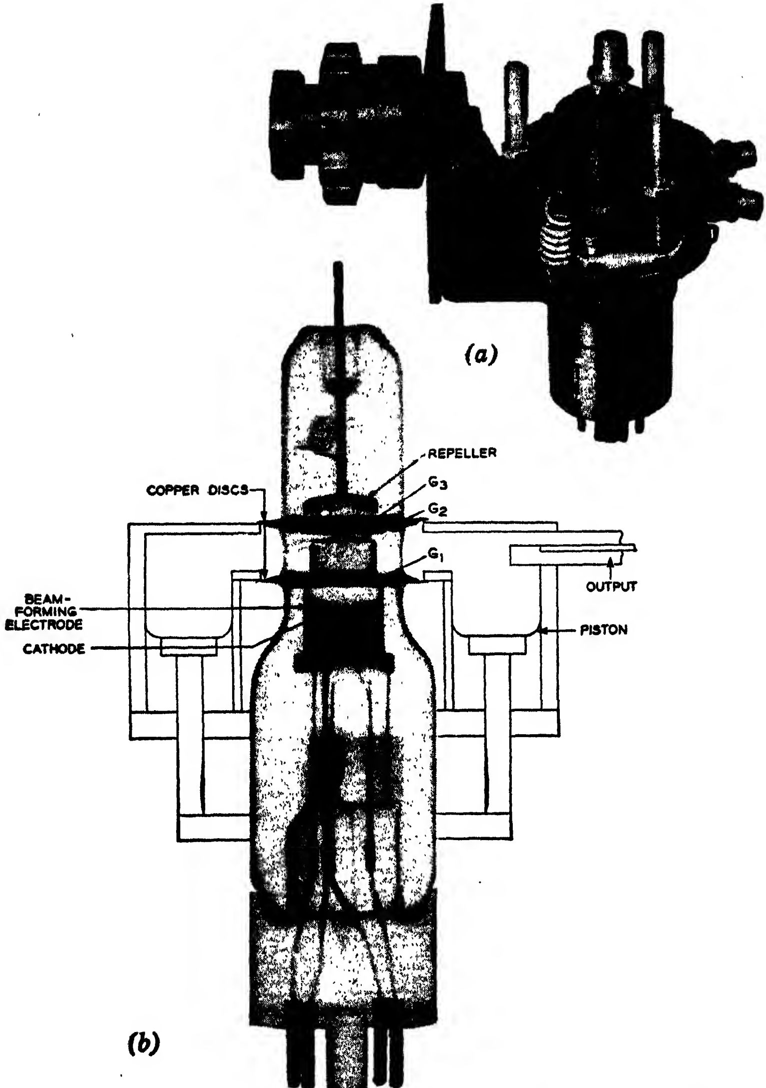


FIG. 3.6 (a) View of Sperry 2K41 klystron tube (courtesy Sperry Gyroscope Company), showing the method of tube mounting with the fine tuning knob on the left. (b) X-ray view of 707 tube (courtesy Bell Telephone Laboratories), showing the inner construction. The external cavity is shown schematically.

A very useful 3-centimeter tube is shown in Fig. 3·5. This is the Shepherd-Pierce tube, or 723A. It will be seen that the cavity is quite small. The tube is tunable, by distortion of the cavity with the tuning screw, over the range from 3.1 to 3.4 centimeters. The output is a small coaxial line which may be used to feed a probe in a waveguide. The shape of the tube lends itself to a very compact mounting on the waveguide installation.

In Fig. 3·6 are shown photographs of a klystron and a McNally tube or 707A.

The tubes shown here were developed primarily for radar purposes and hence are suited to narrow band operation. For wideband operation where tuning is required over a factor of 2 in frequency it is possible to make the electronic section of the tube fit into the inner conductor of a coaxial cavity. This is done in the 2K48 tube described by Clark and Samuel,[4] and such tubes used with coaxial cavities can be tuned over a range from 6600 to 10,700 megacycles. For many purposes in microwave research this ability to tune over a wide range is highly desirable. The operation of the tube is similar to that of the usual reflex klystron. The details of design of cavities and plungers are given fully by Huggins, Zeidler, and Manning.[5]

3·3 HIGH POWER GENERATION OF MICROWAVES: THE MAGNETRON

This section is devoted to the cavity magnetron, or magnetron as it is now almost universally called. It is a remarkable scientific development. Applied to radar it made a versatile weapon which could be said to have made the turning point in the war. Before starting a description of a magnetron, however, it is worth while to consider the factors which must certainly be present in a tube which can deliver very high power. Even in a radar or pulsed accelerator which is "on" for only a small fraction of the time, the ratio of peak power to average power is only about 2000. If the tube delivers 1 megawatt at 50 per cent efficiency the average power to be handled by the tube is 1000 watts, of which 500 watts

[4] J. W. Clark and A. L. Samuel, *Proc. I.R.E.*, **35**, 81 (1947).

[5] In Radio Research Laboratory Staff, *Very High Frequency Techniques*, McGraw-Hill Book Co., 1947, p. 878.

must be dissipated without overheating. If the efficiency is low this figure is increased two or perhaps four times so that the tube must be built to dissipate as much as 2 kilowatts in heat. This will make it very hard to operate a tube having very small spacings, such as we have already described. In addition, the r-f energy must be drawn from some source, which is ultimately a cathode emitting electrons. To get adequate currents of electrons the cathode must be large and rugged.

The science of r-f engineering recognizes all these factors, and they are taken care of in usual radio design. However, it becomes very difficult to achieve all the requirements as the frequency goes into the thousands of megacycles if conventional triode oscillators are used. Remarkable achievements have been made with ring oscillators using several triodes, but these techniques cannot carry over to the centimeter region. The magnetron, which operates on a totally different principle, makes possible a large cathode and good cooling, and has revolutionized radar design.

Cavity Magnetron Oscillators

A cavity magnetron oscillator is shown schematically in Fig. 3·7. The original design was worked out under the direction of M. L. Oliphant at Birmingham University in England. The wartime secrecy imposed on this remarkable tube was such that it is difficult today to give proper credit to those who contributed to its development. The original description was made by Randall, Boot, and Duke in 1941. A very active and ingenious group of whom Sayer was a major contributor kept well in the forefront in England. In this country Clogston, Rieke, Stout, and Young, working under the direction of Collins in the magnetron group of the Radiation Laboratory, have made one series of valuable contributions. Actively collaborating with this group was P. L. Spencer at the Raytheon Company. The group under J. Fisk at the Bell Telephone Laboratories made another series of vital developments, and the very short wave researches have mostly been carried out at Columbia University under J. M. B. Kellogg. Theory has lagged behind experiment in magnetron work. The most extensive studies have been made by Hartree and Stoner in England and by Slater in the United States. Much remains to be accomplished in magnetron design before a complete story of the tube can be given.

A quick account of the action of the magnetron can be obtained by referring to Fig. 3·7. The cathode is a cylinder placed at the axis of a cylindrical copper block having eight cavities consisting of cylinders and slots. In pulsed operation the cathode delivers up to 50 amperes, so that there is exceedingly large space charge around it. When a very high negative voltage pulse is applied to the cathode the electrons are driven to the anode. They are, however, constrained, by a magnetic field perpendicular to their

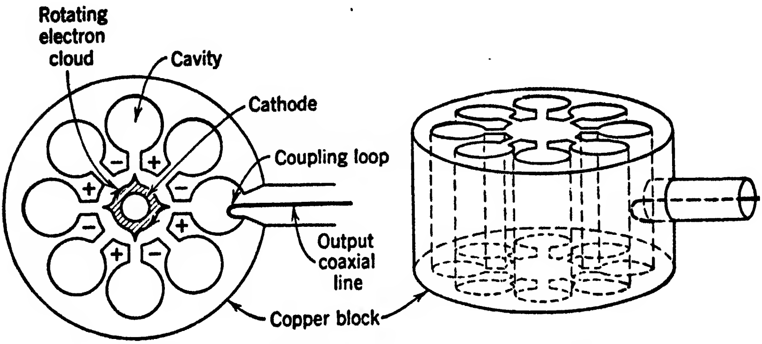


FIG. 3·7 Schematic drawing of a cavity magnetron. A series of 8 resonant cavities opens on to a space containing a cathode. Under the combined influence of the oscillating electric field due to the cavities and a magnetic field parallel to the axis of the cathode, the electrons emitted from the cathode form a rotating cloud which serves to excite further oscillations.

motion, to move in orbits which are a composite of a circle and a spiral. This lateral motion sweeps the electrons past the slots of the cavities, producing oscillations under favorably timed conditions. Since the cavities are of highly conducting material and have low loss, the voltage developed across each gap becomes exceedingly high and in turn conditions the motion of the electrons. The analysis of the motion of the electrons is complex, but the physical process giving rise to oscillations can be thought of by supposing that the broad average of the electron distribution becomes like the spokes of a wheel. This wheel rotates with an angular speed depending on the d-c voltage between cathode and anode, the r-f voltage at each gap, and the magnetic field. If this speed of rotation is such that the electrons are retarded in speed as they are swept across a cavity gap they lose energy to the r-f field and build up the r-f oscillations.

This simple picture is all that can be given without very involved and unsatisfactory mathematical accounts. It suffices to bring out the virtues and defects of the magnetron. In the first place the spacings are all relatively large. A large cathode can therefore be used; in fact the electron cloud will not have the appropriate shape unless the cathode is of large diameter. This guarantees the necessary high currents. In the second place the oscillating elements are made of copper in a block so that heat dissipation is at its best. There is no great problem about putting a high d-c voltage between the cathode and the anode if the vacuum is high. The modern technique of glass-to-metal seals is such that it is not difficult to seal in the two cathode leads and the output coaxial line. The tube should therefore produce quantities of r-f at high efficiency, and it does. Efficiencies of 70 per cent have been reached; a tube is defective if it cannot be operated above 50 per cent efficiency at ten centimeters.

The disadvantages are equally clear. It is a self-excited oscillator and so the rest of the equipment must be designed to consider its whims. Eight cavities are excited by the electron wheel. Such a multiple oscillating system always has more than one mode of composite oscillation, and they will differ in frequency from each other. The distribution of the density in the spokes of the electron wheel may well permit two or more of these oscillations to take place at once, and they will interfere seriously with many operations. In addition, the external system represented by the coupling loop and the coaxial line is not buffered in any way from the oscillations, and indeed it is quite true that the external r-f line is really part of the whole oscillating system. As a consequence reflected power from the line will affect the frequency of the magnetron and may cause a change in mode.

This effect of reflected power in the line on the frequency of oscillation is called *pulling*; the tendency to oscillate in more than one mode is called *double moding*. Pulling and double moding are the two main troubles in magnetron technique.

To continue the description of the cavity magnetron: the block of cavities is enclosed in a cylindrical shell which carries the leads to the cathode. The complete assembly is shown schematically in Fig. 3·8. The whole is placed between the pole pieces of a permanent magnet having a field of around 1000 gauss at 10 centimeters or 3000 at 3 centimeters. The appropriate negative

voltage pulse is applied from a modulator (see Chapter 5), and the magnetron then delivers r-f power.

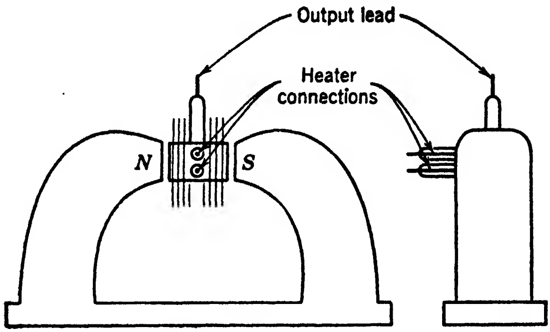


FIG. 3·8 Magnetron and permanent magnet. The magnetron is shown between the pole pieces of a permanent magnet. The cathode is connected through the two heaters leading to the modulator, and the magnetron itself is at ground potential. The output lead, here shown of a primitive kind, goes to the requisite microwave transformer for feeding coaxial line or guide.

Magnetron Scaling

Magnetrons are in common use at wavelengths from 30 to 1 centimeter. Each tube must be specially made for each frequency band, and it is therefore necessary to know whether any simple relations hold to enable the results found at one wavelength to be used at another. Some such relations hold. For example $\lambda H/10{,}600$ is constant, where λ is in centimeters, H in gauss. Thus at 10 centimeters a field of 1300 gauss is used in one successful type of magnetron. The relation shows that a field of 430 gauss would be used for a magnetron of similar design at 30 centimeters or of 4300 at 3 centimeters.

Graphical Description of Magnetron Operation

For the complete description of magnetron performance at least seven variables are required.[6] Those which fit most naturally into experimental r-f work are: H, the magnetic field, V_{dc} the input

[6] This paragraph is based largely on a series of reports by Rieke and those working with him. These reports are: F. F. Rieke and J. E. Evans, "R-f Loading of 10 cm Magnetrons," Radiation Laboratory Report 52-2, August

voltage, I_{dc} the input current, f the frequency, P the power output, and two variables for the load. This list of variables is generally handled by two kinds of diagram, one the performance chart, which focuses attention on the behavior of the magnetron

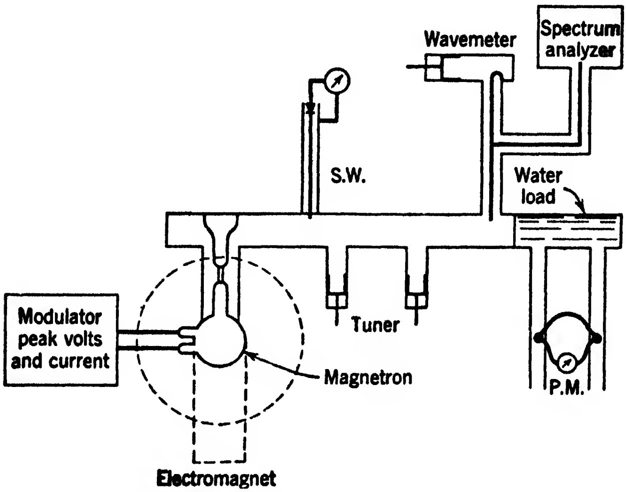


FIG. 3·9 Schematic diagram of equipment for studying magnetrons. S.W. is a slotted section with crystal or thermistor for standing wave measurements, P.M. a power meter to determine the power dissipated in the water load. The modulator should be equipped to read peak volts and current. With this arrangement performance charts and Rieke diagrams can be plotted.

when feeding power into a fixed load, the other the Rieke diagram, which is intended for consideration of the effect of load.

A schematic diagram of equipment which could enable the making of the measurements from which these two diagrams are constructed is given in Fig. 3·9. The magnetron feeds into a line which is terminated in a water load with a thermocouple for measuring the power from the temperature rise and the rate of flow. An electromagnet is used to operate the magnetron so that the magnetic field can be varied. The line is tuned with a double stub tuner, and standing wave measurements are made with a slotted section and a thermistor or crystal. The frequency is

---

24, 1942; F. F. Rieke, "Analysis of Magnetron Performance," Part I, Radiation Laboratory Report 52-10, Sept. 16, 1943; R. Platzman, J. E. Evans, and F. F. Rieke, Part II, Radiation Laboratory Report 451, March 3, 1944.

monitored with a wavemeter and the spectrum with a spectrum analyzer which observes the presence of double modes. The equipment involved in this type of test bench is described in Chapter 4.

The Performance Chart

The general nature of a performance chart is shown in Fig. 3·10. The peak pulse voltage is plotted as ordinate and the peak current as abscissa. It is usual to determine several families of curves by

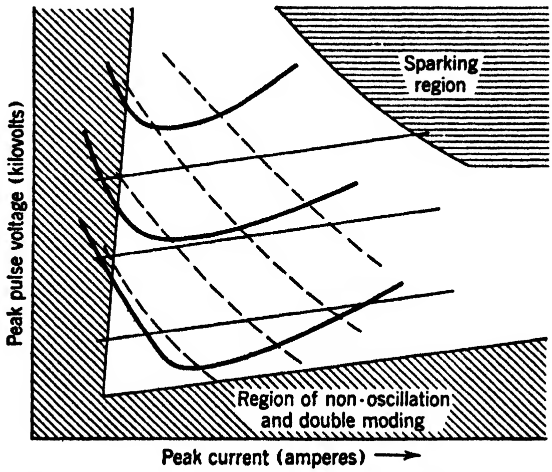


FIG. 3·10 Illustrating the magnetron performance chart. With pulse voltage and peak currents for reference, curves are plotted for constant magnetic field (dashed lines), constant efficiency (thick black lines), and constant power output. The shaded areas represent inadequate operation because of either sparking or improper oscillation.

varying the magnetic field. One series of curves is the simple relation between voltage and current for each value of the magnetic field. A second series is derived by measuring the power output for such variation and plotting the lines of constant power output. A third series is derived by plotting the lines of constant efficiency.

It is found that with a given type of magnetron there are two boundary lines on the performance chart. One is the line beyond which sparking develops. It is in the general area of very high

voltages and currents. The second line marks the region where the tube either will not oscillate or oscillates in more than one mode or in an undesired mode. The performance chart shows these undesirable regions as shaded areas. Within the operable boundaries the chart is of considerable help in verifying that the modulator voltage and the field of the permanent magnet used are actually of appropriate values for good operation at high efficiency.

The Rieke Diagram

A magnetron is not very predictable when it is used to feed power into various kinds of lines. It may be that the line used has a variable impedance (for example an imperfect rotating joint or a variable loading of an accelerator cavity). In this case it is most useful to have some sort of chart which shows how the composite system of magnetron plus variable impedance behaves. This information is provided by the Rieke diagram.

In the plotting of the Rieke diagram a tuner of known characteristics is included in the line to the water load. The magnetic field is kept constant, and the current is brought to the same value at each reading. The results are plotted on a circular diagram which gives the phase and magnitude of the vector which determines the standing wave in the line. Thus for a certain power output which would bring the current to the desired value it will be found that the standing wave ratio is a certain value with the minimum of the pattern at a certain place. The two determining factors, size and phase, of the standing wave pattern are then plotted on the diagram, and the point is labeled with the power. The tuner is then changed and the current again brought to the chosen value. The standing wave constants are again determined and plotted with the new value of the power. By a simple but laborious process a family of constant power lines can be drawn. These are shown as dashed lines for one magnetron in Fig. 3·11. The same process can be applied to frequency, and a second family, shown as continuous lines, can be obtained. Finally, a third family, of constant voltage, can be obtained.

One great use to which such studies have been put is the design of output sections which load the tube properly without obtaining matching at the expense of a high standing wave ratio in the matching transformer.

Notice the effect of a large standing wave ratio on the frequency, the *pulling* of the tube. In the case considered in Fig. 3·11 the pulling varies for different operating conditions. A standing wave

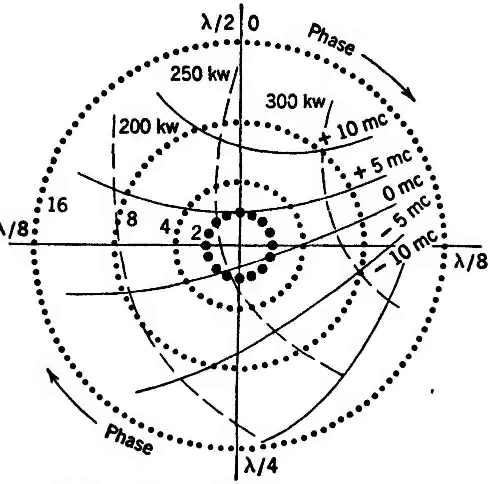


FIG. 3·11 Illustrating the Rieke diagram. This shows the effect of the conditions in the output line on the operation of the magnetron. Change of frequency is shown as continuous lines as the standing wave ratio and phase are varied. The dashed lines show the conditions of constant power output.

ratio of 1.5 can cause a total pulling of about 1 megacycle, which is tolerable.

Modes of Magnetron Oscillation

In Fig. 3·7 a manner of oscillation in which alternate plus and minus occurred between the cavities was indicated. Such a simple mode is by no means the only one. In the first place the cavities can oscillate in higher harmonics so that a 3000-megacycle magnetron can oscillate at 6000, 9000 megacycles, and so on. The amount of power ordinarily generated in one of these harmonics is a very small fraction of that at the fundamental (about one part in 100,000) unless the d-c voltage and magnetic field are specially adjusted to cause the electron wheel to excite these frequencies.

In addition to these widely separated harmonics there exist modes of oscillation at each harmonic which are generally separated only slightly in frequency. These modes have been studied by

Slater[7] and Clogston both theoretically and experimentally. Figure 3·12 is a schematic drawing of the modes of an eight-hole

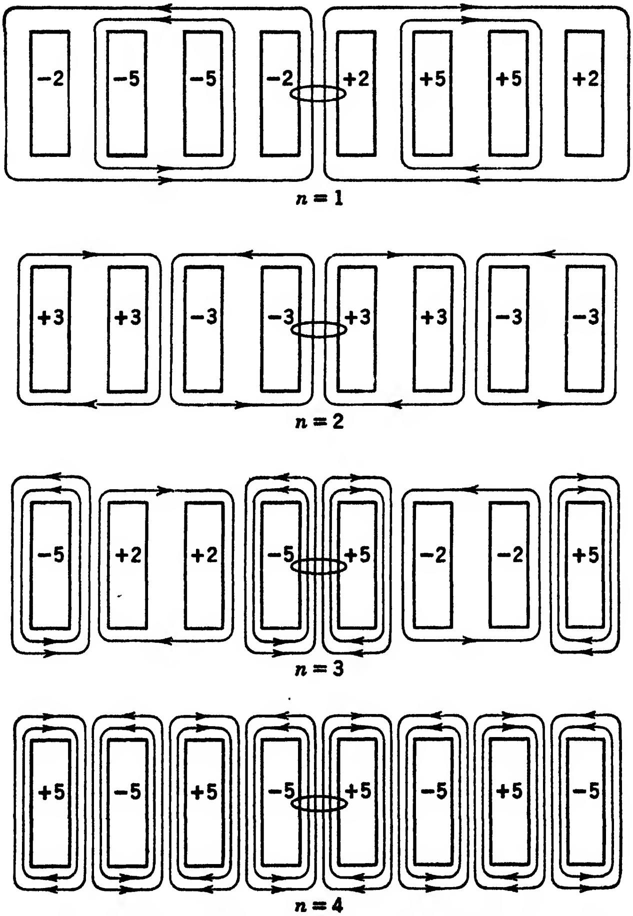


FIG. 3·12 The resonant modes of an eight-hole magnetron. The relative potentials of the metal between the slots are indicated, and the magnetic lines of force are drawn around. The position of the coupling loop is shown at the center of each drawing.

magnetron following a method given by Slater. The modes are characterized by numbers, and it can be seen that the usual mode is the $n = 4$ mode.

[7] J. C. Slater, Radiation Laboratory Report 43-9, August 31, 1942.

For a magnetron having an anode length of 2 centimeters the wavelengths found in one case were

$n = 1$, 9.0 centimeters

$n = 2$, 9.6 centimeters

$n = 3$, 9.8 centimeters

$n = 4$ 9.9 centimeters

The unwanted $n = 3$ mode is thus seen to be quite close in wavelength to the desired $n = 4$ mode. A technique of "strapping" introduced by Spencer [8] and independently by the Birmingham University group greatly alleviates this situation. Connecting by straps the four parts of the copper block which swing in voltage together in the $n = 4$ mode has the effect of spreading the wavelengths of the various modes much farther apart. If the "tightness," meaning the closeness of the straps to the center, is increased, the wavelengths of the four modes move apart. A magnetron so strapped would thus operate at longer wavelengths but the unwanted modes would be very much farther away in wavelength and therefore considerably harder to excite.

Another very important method of mode separation is employed in the "rising sun" magnetron developed at Columbia University Radiation Laboratory. Here cavities of alternately different sizes are used. The cavities are actually slotted in form, and one set of slots is about 75 per cent longer than the other. The individual frequencies of the slots are thus quite different. The result is that the modes, instead of being in one gradually increasing set of wavelengths as for an unstrapped magnetron, are in two groups, one of longer wavelength and one of shorter. The separation of modes is nearly as great as for heavy strapping without the use of straps at all. The advantage of this design is that it permits the use of larger numbers of cavities, which is advantageous at short wavelengths as the rotating electron cloud cannot rotate fast enough to excite very high frequency oscillations if it has to describe too large an angle. Accordingly the rising sun design is found in magnetrons intended for use in the 1-centimeter region.

[8] P. L. Spencer, U. S. Pat. 2,408,234; 2,408,235; 2,408,237.

Tunable Magnetrons

In all phases of microwave design, research has gone into means for obtaining broad tuning. This is also true of magnetrons. The problem in the case of magnetrons is more acute on account of the multiplicity of cavities. Several solutions have been reached which are reasonably satisfactory and have gone into production. The method employed introduces metal into the resonant cavities either in the form of a slug in the circular inductive part, or a vane in the narrow capacitive part. The tuning mechanism is enclosed in the vacuum by means of either a flexible metal diaphragm or a long bellows. A flat gear train is provided to operate the tuning. The output power is never quite so high as the full power possible, but it is not greatly reduced. The 2J21 can be tuned over a range from 8500 to 9600 megacycles, the 5J26 from 1220 to 1350 megacycles.

Practical Design Features of Magnetrons

Of great importance to magnetron design is the method of matching to the line. The magnetron itself seems to be capable of developing all the power that can be handled. At these high powers the simple technique of removing the power from a loop to a tungsten rod of perhaps 1/8 inch in diameter and then feeding a coaxial line results in breakdown due to high field far below the available power. The authors remember much ozone resulting from this procedure. It is therefore necessary to consider carefully the means of getting the power into the desired line. The magnetrons operating in the 1-centimeter region have the power taken directly into waveguide from a slot in the back of one of the cavities. High power magnetrons at 10 centimeters, like the HK-7, use the loop method but combine it with a taper transformer to a 3/4-inch inner conductor of a 2 1/2-inch coaxial line, the transformer being within the vacuum. The design problem of the output transformer is not simply that of tapering to a big diameter inside the vacuum seal but is also vitally concerned with the amount of pulling.

A second important consideration is the cathode structure. The problem of heat dissipation in the anode has to all intents and purposes been solved by the use of the solid copper block with

cavities cut into it. To attain the powers given by these cavities requires the emission of very high currents, 50 amperes being possible. Modern design uses a sintered-nickel base for oxide coating with a heating coil tightly wound inside.

Magnetrons Commercially Available

The great impetus in magnetron construction came from radar. For this reason the great majority of magnetrons available are intended for pulsed operation at high voltages and high currents

TABLE 3·1 CHARACTERISTICS OF VARIOUS MAGNETRONS

Type	N	Resonator System	λ	Peak Volts	Peak Amperes	Gauss	τ	P	Efficiency (per cent)	Pulling Figure
3J21	18	R.S.	1.25	15 kv	15	8000	0.5	60	26	17
725A	12	H. and S.	3.2	13	12	5650	1	56	51	13.5
2J51	12	H. and S.	3.5	14	14	4000	1	60	46	12
4J50	16	H. and S.	3.2	22	27	6900	1	280	66	12
4J45	8	H. and S.	10.7	27	65	2900	1	1100	68	10
4J26	8	H. and S.	24.0	27	46	1400	5	700	60	3
HP10V	10	Vane	10.7	48	140	1800	1	2600	38	10

N stands for the number of resonators.

R.S. means Rising Sun, H. and S. means hole and slot, vane has vanes to separate the cylindrical space into cavities.

τ is the pulse width in microseconds.

P is the power delivered in kilowatts, peak.

The pulling figure is the frequency shift in megacycles for a standing wave voltage ratio of 1.5.

with a low ratio of on time to off time. So many radar-type magnetrons were made that tubes once costing hundreds of dollars are now advertised for a few dollars. In Fig. 3·13 are shown pictures of two magnetrons. Characteristics for a number of tubes are listed in Table 3·1. A few words can be said in comment. Magnetrons have been developed from 800 to 30,000 megacycles. The peak output powers readily available remain at about 1 megawatt,

(a)

(b)

FIG. 3·13 (a) A high power 10-cm magnetron, the 4J33 (courtesy Raytheon Manufacturing Company), cut away to show the cavity structure and strapping. (b) A 3-cm magnetron with magnet attached (courtesy Bell Telephone Laboratories).

though up to 3 can be obtained with the HP10, which is not yet on the market. These figures apply up to about 4000 megacycles. Above this the power drops about as the $\frac{3}{2}$ power of the wavelength, being about 30 kilowatts at 25,000 megacycles. The small size of magnetrons above 10,000 megacycles renders them suitable to be designed in the "packaged" form, having the magnet integral with the tube. Generally the voltage required for operation is high, in excess of 10 kilovolts, but special low voltage magnetrons were developed for lightweight radars which operate at two or three thousand volts. The efficiency gets higher as the voltage of operation is increased.

Magnetrons can be used as generators of continuous power. These were not required in such large numbers during the war, and so they are not so commonplace. The power output of a continuously operating magnetron is about the same as the average power of a pulsed magnetron, though design changes have to be made to ensure this, since pulsed magnetrons operate at high voltages, and continuously operated magnetrons will not last long under such conditions. Power of the order of a kilowatt can be obtained.

Summary

To summarize this chapter it can be said that the development of cavity oscillations for bench purposes can be achieved by triodes and velocity-modulated tubes. The latter can probably be made to operate at as short a wavelength as 0.2 centimeter. Powers up to 3 megawatts in pulsed form at 9 centimeters and longer, or 30 kilowatts at 1 centimeter, can be obtained with magnetrons. These developments are briefly sketched.

ADDITIONAL REFERENCES

H. D. Hagstrum, "The Generation of Centimeter Waves," *Proc. I.R.E.*, **35**, 548 (1947). A very good general account.

E. D. McArthur, "Disc Seal Tubes," *Electronics*, **18**, 98–102, Feb. 1945; A. M. Gurewitsch and J. R. Whinnery, *Proc. I.R.E.*, **35**, 462 (1947). These two articles give full information on lighthouse tubes.

J. R. Pierce and W. G. Shepherd, "Reflex Oscillators," *Bell System Tech. J.*, **26**, 460 (1947).

J. Fisk, H. D. Hagstrum, and P. L. Hartman, *Bell System Tech. J.*, **25**, 167 (1946). A very readable account of magnetrons.

CHAPTER 4

MICROWAVE TECHNIQUE

The development of microwave technique in the past six years has largely centered around radar. It was mentioned in Chapter 1 that a new development is of importance by reason of the extremes it renders attainable. Microwaves offered narrow beams without excessively large equipment, with no lessening of available power. This was a tremendous step forward for airborne radar, about which microwave radar originally developed, and proved to be a step forward in all radar. At the end of the war every possible radar job was being done by microwave equipment.

The purpose of this chapter is to outline microwave technique as it is used today. It cannot possibly be done at any length in a single chapter, but enough can be given to enable the reader to gain an understanding of the part microwave technique is likely to play in future research.

The development of microwave manipulation from a temperamental semi-miraculous process to a smooth and reliable operation is the work of many groups of individuals. Perhaps the most credit should go to several British organizations, notably the Telecommunications Research Establishment at Malvern. Active work was pursued in the United States and Canada by such strong research organizations as the Bell Telephone Laboratories, General Electric, and Sperry Gyroscope Company. In the center of all this activity was the Radiation Laboratory of the Massachusetts

Institute of Technology, and it is largely the Radiation Laboratory technique which is outlined here.[1]

4·1 DETECTION AND MEASUREMENT OF MICROWAVES

The first problem the neophyte encounters when he starts work with microwaves is that of convincing himself that he actually has microwaves in his equipment. Those who work with magnetrons get to smell ozone and light cigarettes at thousands of megacycles, but the ordinary worker on radio-frequency has to use subtler means. The simplest of the means of detecting microwave power is a crystal.

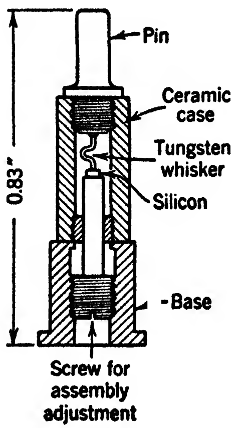


FIG. 4·1 Schematic drawing of a 1N21B crystal.

Crystal Detection

Crystals are usually mounted in a special cartridge, one of which is illustrated in Fig. 4·1. It consists of a tungsten whisker held in point contact with a silicon crystal in as standardized a manner as manufacture will permit. Much research has been done on crystals to improve their efficiency as frequency converters and their ruggedness in the sense of withstanding short heavy shocks of high power. Since the crystal has to be part of the r-f system it is necessary that the cartridge be so designed that the impedance of different

[1] As general reference material we have relied on Radiation Laboratory Reports and the Sperry handbook, *Microwave Transmission Design Data.* The reference material used in general is listed here; other references are given as they occur. (*a*) Sarbacher and Edson, *Hyper- and Ultrahigh Frequency Engineering,* John Wiley and Sons, Inc. 1943; (*b*) J. C. Slater, *Microwave Transmission,* McGraw-Hill Book Co., 1942; (*c*) W. W. Hansen, lecture notes; (*d*) "Microwave Transmission Design Data," Sperry Gyroscope Company (1944); (*e*) H. Krutter, "Explanation of Impedance Matching," Radiation Laboratory Report T-6; (*f*) S. Goudsmit, "Reflection Coefficients and Impedance Charts," Radiation Laboratory Report T-11; (*g*) S. Seely, Radiation Laboratory Report T-10; (*h*) Radiation Laboratory Report T-13, prepared by the R-F Group.

crystals is the same in order to avoid new impedance matching each time a crystal is changed. The two most successful types of crystal, the 1N21B and 1N23B, operate in the 10- and 3-centimeter regions respectively. The rectifying material is silicon with $\frac{1}{10}$ per cent "dope," and the point a sharpened tungsten wire which should make as small a surface of contact as is feasible. It should be borne in mind that the crystal should have high "burnout" power.

The resistance of a crystal as measured by an ohmmeter of the usual type differs according to the direction of the voltage used in the measurement. The ratio of the two values is often called the front-to-back ratio, although it is by no means a fixed quantity because it depends on the applied voltage. Burnout is generally indicated by a big drop in the front-to-back ratio which can change from around 20 to around unity. A small meter set up for measuring quickly this front-to-back ratio is a comforting piece of equipment since failure to get a meter reading is one of the troubles of r-f work, and it is pleasant to know that the crystal ought to work. Burnout occurs at from 1 to 10 watts for rugged crystals. It is wise to subject them to less.

The use of a crystal for r-f detection can be illustrated by a description of the process for getting a klystron to oscillate at the right frequency. For this one needs a klystron and power supply, a wavemeter (consisting of a coaxial line cavity with plunger to vary its length, see page 90), a length of cable, the right connectors, and an appropriate crystal holder. The apparatus can be set up as in Fig. 4·2. The klystron pickup loop is connected by the cable to the tee-junction which leads to the wavemeter on the one hand and the crystal on the other. The milliammeter is connected across the crystal, and it reads the rectified current. The distance from the actual junction of the inner conductors to the wavemeter should be about one-half wavelength, for then the impedance at the tee is approximately that at the wavemeter.

Power is then applied to the klystron, and the odds are that it does not oscillate. The reflector plate potential is changed until the milliammeter deflects. This deflection shows that the klystron is giving r-f power. The position of the plunger in the wavemeter is then changed gradually until the crystal current shows a definite dip. No extra rise should precede or follow the dip, as that is an indication that the wavemeter is too tightly coupled.

The coupling can be reduced by rotating the loop in the wavemeter by a slight turn on the input. The wavemeter reading then

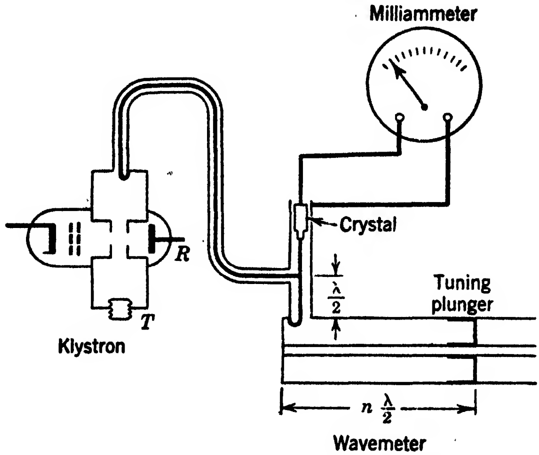


FIG. 4·2 Illustrating the method of connecting a klystron to a wavemeter and crystal detector to measure wavelength and set the oscillator on frequency.

indicates the wavelength, either directly or as the difference between two positions of dip. The coarse tuning screw T on the klystron (which may take one of several forms according to design) is then turned, with appropriate adjustment of the reflector voltage to keep the klystron oscillating, until the wavemeter reads the required wavelength.

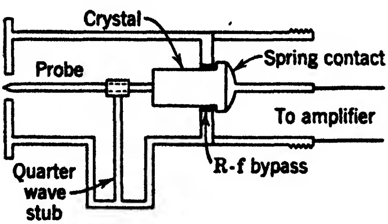


FIG. 4·3 Showing the construction of a probe suitable for standing wave measurements. The crystal detector is placed in a ring of dielectric which serves as a bypass for the microwave component. The rectified component is picked up by an amplifier.

It is often desired to make measurements with a crystal on a *pulsed* r-f circuit. In such a case it is convenient to use the rectified current to develop a voltage in a resistance in series with the crystal and to amplify this voltage with an audio-amplifier set to maximum sensitivity at about the pulse repetition frequency. Very simple amplifiers will suffice for this.

For the measurement of standing waves in a slotted section of waveguide or coaxial line a probe such as that illustrated in Fig. 4·3 is used. Note the use of a quarter-wave stub to support the line going to the crystal, and the r-f bypass consisting of a ring of thin insulator around the case of the crystal. For high power measurements the probe must be very loosely coupled to the line. This adjustment is made by sliding the whole center assembly in or out as needed.

A crystal gives a current approximately proportional to the r-f power at the probe. This is in turn proportional to the square of the r-f field at the probe. The accuracy of this relation is fair and is enough for many r-f measurements.

Bolometers and Thermistors

Although the crystal is a most useful means of detecting and measuring microwave power it lacks any real show of precision. In order to have available an instrument which is capable of measuring power as a definite number the use of bolometers and thermistors has become common. Both operate on the same principle: the bolometer is a fine wire (for example a Littelfuse), the resistance of which increases when heated by any means, including r-f power; the thermistor is a bead of semiconducting material with a very large *negative* temperature coefficient of resistance.[2] The resistance changes by about 4 per cent per degree. The change of resistance is employed to change the balance of a Wheatstone bridge, which can be restored to balance by a known amount of d-c power, which is then the measure of the r-f power; or the unbalance current can be calibrated in terms of power and read directly.

Thermistors seem to be preferable in practice. They are able to take greater overloads without burning out, and this is a very desirable characteristic if the system under test is at high power. The thermistor is a very small bead held by two fine wires supported by a glass envelope if glass does not introduce too much loss (i.e., down to 3 centimeters wavelength). Resistances are of the order of 100 ohms when operating in the bridge. The thermistor has to be placed in a properly matched line, which can be of

[2] See J. A. Becker, C. B. Green, and G. L. Pearson, *Bell System Tech. J.*, **26**, 170 (1947).

various designs. Matching by means of a tapered section seems to be very satisfactory.

Wavemeters

In principle any cavity having high selectivity (Q) which oscillates in a known mode can be used as a wavemeter. Actually a coaxial line cavity is very convenient, since it has a definite mode of oscillation for which the wavelength is the same as in air, and it is mechanically convenient from the point of view of moving a shorting plate at one end. The principle of operation of a coaxial line wavemeter can be seen in Fig. 4·2. The problem of a perfect sealing of the end of the line is not quite simple. If sliding contact is made at the exact end the currents through the contact are at maximum, and any poor contact will result in imperfect closure of the cavity and reduce the Q. Therefore the sliding plunger is made re-entrant, with the fingers which make contact somewhere near a quarter wavelength out. In this position the current at contact is low, the voltage high, and therefore a poor contact has relatively little effect on the Q. Doing this has the effect of sacrificing the absolute nature of the reading unless two minima at different lengths can be found. The difference between these lengths is accurately a half wavelength. This procedure is generally adopted.

If a higher Q than is given by a coaxial cavity is needed for very precise wavelength measurement a cylindrical cavity wavemeter can be used. It needs to be calibrated. For calibration purposes standard cylindrical cavities of known length and diameter can be made.

An example of the use of a wavemeter has been given earlier in the chapter. It is possible to use the wavemeter in series with the line. It is a little simpler to connect this way, but it has the disadvantage that there is no reading until the wavemeter is close to the right wavelength.

Attenuators

A surprisingly important piece of equipment for microwave work is a calibrated attenuator, because widely different power levels are encountered in practice and the simple comparison be-

tween two power levels may be most informative. Attenuators for measurement work are commonly based on the principle of propagation in a waveguide beyond cutoff. For propagation down a cylinder of radius a in the TE_{11} mode the attenuation in decibels per centimeter, α, is given by

$$\alpha = 8.7\left[\left(\frac{2.405}{a}\right)^2 - \left(\frac{2\pi}{\lambda}\right)^2\right]^{1/2} \tag{4·1}$$

where a is the radius and λ the wavelength. It can be seen that a wide range of attenuation is easily obtained if the radius is small

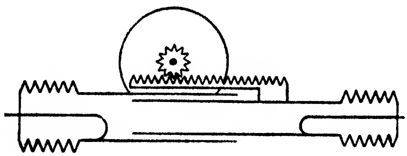


FIG. 4·4 Illustrating the principle of an attenuator for making known reductions in r-f power. The two loops couple in to an adjustable length of waveguide which is far less than the cutoff diameter. The attenuation is theoretically calculable.

compared to the wavelength. A schematic drawing of an attenuator of this type is shown in Fig. 4·4. There is always an "insertion loss," which is about 15 decibels, after which the attenuator has an accurately linear calibration.

Another kind of attenuator is frequently used to dissipate power. For this purpose a semi-conducting center conductor in a coaxial line can be used. In a waveguide, semi-conducting strips can be placed at an angle to eliminate reflections and guarantee absorption. Another termination for absorption consists of powdered iron suspended in an insulating medium (this is called Polyiron). In addition a water termination is often used for power measurement purposes. To dissipate high power a sand load is used.

Spectrum Analyzers

An exceedingly powerful and versatile tool for the study of any frequency-dependent microwave quantity was developed by J. L. Lawson and his group at the Radiation Laboratory and is known as the spectrum analyzer. This instrument makes use of

the electrical tuning properties of a reflex klystron and combines a sweep voltage on an oscilloscope with a synchronized sawtooth voltage on the reflector plate of the klystron. The horizontal deflection of the oscilloscope is then reasonably proportional to the change in frequency from the midpoint frequency of the klystron. If some frequency-dependent quantity can be detected and amplified sufficiently to give a vertical deflection, it is immediately apparent which frequency or range of frequencies is responsible for the effect.

A rudimentary spectrum analyzer can be made by developing a sawtooth voltage with enough power behind it to drive the reflector plate of a klystron and sufficient amplitude to give a horizontal sweep on an oscilloscope. Spectrum analyzers for use in radar techniques are more elaborate. They are employed to study the frequency spectrum of pulsed magnetrons both on the test bench and in operating radars. To do this in one unit requires the following components.

1. The swept frequency klystron as described above.
2. A crystal mixer to convert the input signal plus the output of the klystron to 30 megacycles, or whatever intermediate frequency was chosen.
3. A narrow band amplifier tuned to the intermediate frequency to pick one component of the frequency spectrum.
4. A calibrated attenuator to handle a range of power levels and enable relative power measurements to be made.
5. A second oscillator and wavemeter to enable the actual frequency of the picture presented on the oscilloscope to be read.

In addition, circuitry for providing sweeps is needed. A block diagram of a spectrum analyzer is given in Fig. 4·5 (*a*), and in Fig. 4·5 (*b*) is shown schematically the appearance of the frequency envelope of a pulse of microwave power. As each pulse from the power source beats with the slowly varying frequency of the swept klystron, a variable amount of power occurs at the frequency of the narrow band amplifier. The appearance on the screen is then that of a series of pulses whose amplitude fits the envelope of the frequency distribution of the source. The rate of sweeping of the klystron must be much slower than the repetition rate of the pulsed oscillator. A rate of about 25 cycles is satisfactory.

With a 30-megacycle intermediate frequency (i-f) amplifier there are naturally two possible positions for the spectrum since the i-f amplifier does not know whether the klystron frequency is above or below that of the oscillator under test. This fact is

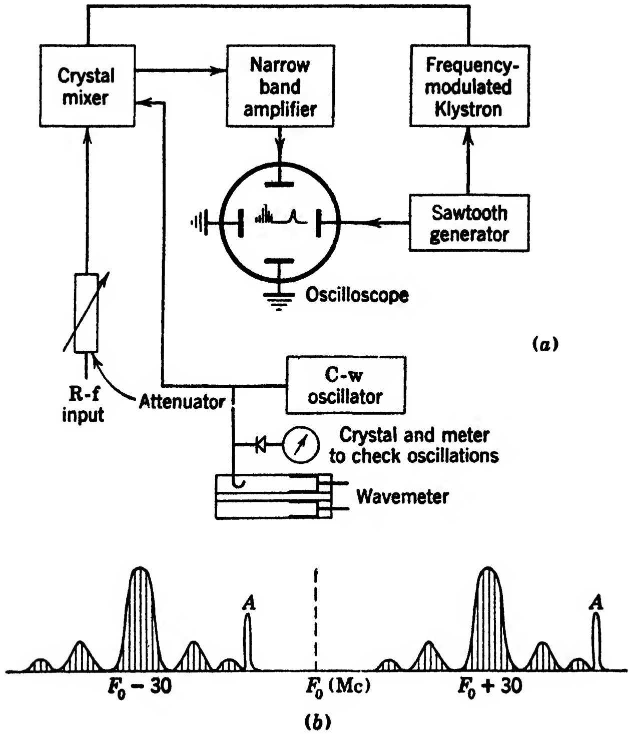


Fig. 4·5 (*a*) shows the block diagram of a spectrum analyzer, an instrument which uses a frequency-modulated klystron and oscilloscope to present the results of a varied frequency on an r-f system; (*b*) shows how a magnetron and a c-w reference frequency *A* have two spectra produced by beating with the local oscillator.

shown in Fig. 4·5 (*b*). In the same diagram the appearance of a continuous oscillation 10 megacycles too high is shown. This is often used as a calibrating device.

The appearance shown in the figure is characteristic of the spectrum of a pulse of radiant energy, which is discussed fully elsewhere, notably Chapter 5.

In the art of radar one fundamental problem encountered is that of luring microwave power out of a magnetron, into a line, onto the surface of a reflecting paraboloid, and back down again, if a target happens to give any return power. It is quite easy to avoid this completely at any one of the stages, as those who have tuned radars in the field know quite well.

There is nothing mystic about microwaves. There do, however, exist waves in two directions as a general phenomenon. In radar particularly, and also elsewhere, it is of importance to minimize the amplitude of one of the two waves. This process is known by the name of "impedance matching." In the early days of microwave radar, when the performance of a set was a function of the skill of the operator (there being every tuning adjustment imaginable), the process of impedance matching was regarded as a black art. This it still is, if any equipment such as bends, tees, dipoles, and paraboloids is hooked together without previous bench testing and no monitoring equipment, and radiofrequency is turned into it. Microwave radar rapidly grew out of this stage, and for the last two years has had almost no tuning adjustments at all for the operator to handle.

This state of affairs was reached by replacing cut-and-try attempts at "tuning" with careful bench measurements. This will have to be done, in limited degree, in any experimental research or demonstration equipment that is expected to operate reliably. It involves a simple, rather pleasant, doctrine, which is here briefly described.

Reflection Coefficients and Standing Waves

Any transmission line which is intended to be simple and is infinitely long has one mode of power transfer and no reflected wave. As soon as anything occurs to change this the above statement ceases to be true. Power is reflected if the line is terminated, in general, or if any obstacle is introduced into the line, or if the line is broken in whole or in part. Inherently nothing can be done about this fact. But it is possible to measure quantitatively the reflection coefficient, examine the process causing the reflection,

and then add other obstacles or changes in such a way as to reduce to a minimum the reflection in the part of the line involved. The first job to do therefore is to find some way to measure all that can be measured about a reflection coefficient due to some unavoidable change in the line. This is done by standing wave measurements. The second job is to examine the possibility of eliminating the return wave by a matching device without introducing some other complication (like breakdown in the line due to excessive fields). It is generally found that a heavy reflection cannot usefully be eliminated. It is wise to return to the basic design and work on some radical change rather than to rely on the process of impedance matching to save the situation. However, where a moderate match is initially present it is possible to improve it considerably by transformers; this process will now be considered.

The measurement of standing waves is a relatively simple matter. It can be done with a slotted section of line or waveguide and a probe with a crystal detector, or with a thermistor bridge. The results are generally expressed in the form of a voltage-standing wave ratio, meaning the ratio between the voltage across the line at maximum to the voltage at minimum. The square of this ratio is the power-standing wave ratio. The position of a minimum on the line is also of importance, as we shall show. Most bench measurements therefore yield the voltage-standing wave ratio, or simply standing wave ratio, and a minimum position.

Standing waves are due to the superposition of waves traveling in opposite directions. We have already called attention to the fact that a wave can be described in terms of a component of a rotating radius arm whose motion is determined by both time and position. Now consider one point on an r-f line. It can be in coaxial line or waveguide or even the space near an antenna pointing at a reflector. Since we do not change the position on the line the motion of the radius arm is determined only by time. We know that it rotates with an angular velocity ω related to the frequency f by $\omega = 2\pi f$. Now suppose that there is an outgoing wave of amplitude *unity*, and a reflected wave of amplitude a, with some phase relation to the outgoing wave given by the angle α. The two are at the same frequency so that, at the point under consideration, the two radius arms will be added vectorially and the resultant will rotate, producing the required sine wave variations in either an x or a y component. If the phase is such that

the two arms add there will be a large amplitude, and if the two arms subtract there will be a small amplitude.

Now move the point under consideration back toward the generator. The phase of the outgoing wave will go backwards, that of the reflected wave will increase. The amplitude will therefore vary as the point goes back toward the generator, passing through maxima and minima. The rate of variation will be such that each *half* wavelength along the line will cause a *relative* phase shift of 360°, and so the variation in voltage will have this periodicity in distance.

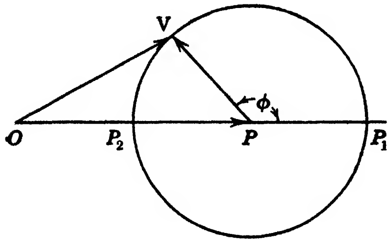


FIG. 4·6 The vector addition of the incident wave, amplitude OP, and the reflected wave, amplitude PV. The measurement of the standing wave ratio gives OP_1/OP_2 from which the amplitude of the reflected wave can be found. Measurement of the phase angle ϕ can be related to the nature of the reflecting process in a systematic way.

The variation of the *amplitude* of the electrical oscillations as the point considered is moved toward the generator is shown in Fig. 4·6. OP is the vector for the incident wave. PV is the vector for the reflected wave. In this diagram we keep OP fixed and refer PV to it. The rotation of PV is therefore a composite of the angular retardation of OP and the advance of PV. The resultant vector which describes the time variation at any point is therefore OV, which varies in amplitude between OP_1 and OP_2. The distance traveled along the line while PV goes all the way round the circle is *one-half* wavelength.

We can now see what we do when we measure standing waves. Moving the probe of the measuring apparatus along a slot corresponds to exploring the variation of OV. The maximum and minimum values tell us the values of OP_1 and OP_2. These values enable the amplitude of the reflected wave, namely one-half P_1P_2, to be measured. For this reason 90 per cent of standing wave measurements are simply aimed at determining the standing wave ratio, since the reflected power is so often the desired quantity.

For example, suppose r is the voltage standing wave ratio; unity and a the amplitudes of the incident and reflected waves. Then

$$r = \frac{1 + a}{1 - a} \qquad a = \frac{r - 1}{r + 1} \tag{4·2}$$

Suppose we register a value of 100 for r before a line is filled with gas and 10 after it is filled. Then before filling we have a very nearly 0.98, whereas after filling it is close to 0.82. The ratio of the amplitude after gas filling to that without gas is therefore 0.82/0.98 or roughly 83 per cent. This illustration gives a measure of gas attenuation and shows how the standing wave ratio, with no profound analysis, leads to direct measurements of reflected amplitudes.

It is clear that more information is available which is not used in so slight an analysis as the above. The reflection is due to a specific cause, and we have treated it as perfectly general. More information can be gained from the phase of the reflected wave, which can be measured by observing the position of the minimum of the standing wave pattern. Such measurements could be made and correlated with known types of reflection and a systematic science built up which would enable corrective measures to be devised intelligently. This indeed has been done. However, it has not been done without some previous ideas as guide. A considerable body of information existed before microwaves were developed, information collected in the study of transmission lines used at relatively low frequencies. Therefore, by common consent, the simple field theory of microwaves has been merged with the line theory of radio engineering, and the terminology used is taken from the latter. This brings with it some surprises, as for example the fact that a change in the radius of the inner conductor of a coaxial line is a "transformer," and a narrowing of a waveguide is an "inductive" iris. It seems to work, and the fact that during the war many physicists adopted it cheerfully argues that it is effective terminology. It means we need to translate the language of fields into voltages and currents.

Voltage and Current Ratios when Reflections Occur: Impedance Variations

Consider the voltage and current ratios in a line when waves are traveling in it in both directions. We stress here the point we have already made that in a reflected wave the magnetic field is

reversed with respect to the electric field. Since the magnetic field is produced by the current, the current and voltage in the reverse wave are in the opposite sense from the direct wave. Thus for a wave traveling forward the voltage will be directed outward for current flowing forward along the center conductor. For a wave traveling backwards the voltage will be directed inward for current flowing in the same direction.

On page 34 we showed how a knowledge of the field in a coaxial line could be used to calculate the ratio of the voltage across the line to the current flowing in the line, a ratio known as the characteristic impedance. When reflections are present both the voltage and the current will be variable at different points of the standing wave pattern, and accordingly there will not be a constant impedance but one which is periodically varying. Since it is proposed to consider coaxial line as a form of transmission line it is necessary to know the nature of the impedance variation.

Some information about this variation can be gained by a simple mathematical analysis. Represent the outgoing current by $i = e^{j(\omega t - kx)}$ and the outgoing voltage by $Z_0 e^{j(\omega t - kx)}$. Since the ratio of voltage to current is defined as impedance we have just assigned the line an impedance Z_0. Now suppose that on reflection both voltage and current suffer a phase shift of ϕ radians. Suppose also that there is a reduction of amplitude by a factor a. We have just pointed out in the previous paragraph that in the reflected wave the voltage is of opposite sign to the current; therefore for the reflected wave we have $i_r = aZ_0 e^{j(\omega t + kx + \phi)}$ and $V_r = -aZ_0 e^{j(\omega t + kx + \phi)}$. Note that in the reversed wave the sign of kx is positive, denoting progress in the negative x direction. Adding the two we get for the current and voltage i_s and V_s

$$i_s = e^{j(\omega t - kx)} + ae^{j(\omega t + kx + \phi)}$$

$$V_s = Z_0 e^{j(\omega t - kx)} - aZ_0 e^{j(\omega t + kx + \phi)}$$

Collecting terms we get

$$i_s = e^{j\omega t}[e^{-jkx} + ae^{j(kx + \phi)}]$$

$$V_s = Z_0 e^{j\omega t}[e^{-jkx} - ae^{j(kx + \phi)}]$$

The ratio of these, which is the *impedance* if the voltage is in the numerator or the *admittance* if the current is in the numerator, is

a very simple quantity. Taking the ratio of voltage to current we have for the actual impedance Z as a function of position

$$Z = Z_o \left[\frac{1 + ae^{j(2kx+\phi)}}{1 - ae^{j(2kx+\phi)}} \right] \tag{4·3}$$

The impedance varies along the line due to the reflection. A trivial but important case is that in which Z is the same as Z_o at all points, in which case the attachment is a line of the same impedance. This can only be true for $a = 0$, or no reflection. The problem of elimination of reflected energy can therefore be restated as the problem of attaching the proper impedance at the end of a section of line. Notice that when the exponent is zero we get

$$\frac{Z}{Z_0} = \frac{1 + a}{1 - a} = r \tag{4·4}$$

The voltage standing wave ratio is therefore a direct measure of the ratio of impedances with and without reflections.

The manner in which the composite impedance due to both outgoing and return waves changes is of interest. As x varies, the quantity $ae^{j(2kx+\phi)}$, which in the geometrical representation of Fig. 4·6 is the rotating radius arm PV, also varies. As $2kx$ increases by π the value of $ae^{j(2kx+\phi)}$ reverses sign, while it has the same magnitude. The value of Z/Z_o is therefore the reciprocal of its initial value. Hence in traveling down the line a distance $2kx = \pi$, which is a quarter wavelength, we have moved to a point which has changed a low impedance in a high one. With obvious notation

$$\frac{Z}{Z_0} = \frac{Z_0}{Z_{\lambda/4}}$$

or

$$ZZ_{\lambda/4} = Z_0^2 \tag{4·5}$$

It will be seen later that much use is made of this property of a quarter wave section.

Reducing Reflections: Impedance Matching

If a transmission line, in practice, has considerable reflected power, it is obviously undesirable. Power is wasted, and the

presence of standing waves in the line causes regions of high voltage and may therefore cause breakdown. It is possible to introduce additional changes in the line so as to remove the reflections at the expense of a local region of reflection which is not serious with regard to power loss or breakdown. Such a process is called impedance matching.

Suppose a line is intended to feed an antenna but the feed at the end is found to cause reflection when the standing wave ratio is measured on the bench. Now it is conceivable that a diaphragm, which would also cause a reflection, could be put into the line. This diaphragm does two things of importance. In the first place it produces an added reflection which can perhaps cancel the reflection to be removed. In the second place it confines radiation in the region between it and the antenna feed. This fact can perhaps be used to build up a region of higher field intensity from which, *in spite of the poor antenna feed*, the whole power available from the generator can be radiated. If such a combination is to take place it would be expected that both the size and position of the diaphragm would have to be right.

The diaphragm discussed above is similar to a transformer in that it depends for its action on a region of high field intensity. The more usual transformer has this field in the iron core. In the present case it is in a confined region of line. Diaphragms, sections of line, and added lines are called transformers even though they bear no similarity in appearance to ordinary transformers.

In order to appreciate the process of impedance matching consider how the above antenna feed could be matched to the line. The apparatus is shown in Fig. 4·7. A source of microwave power at the right frequency (and supposedly correctly matched) feeds into the line. A standing wave measurement is then made by moving a loosely coupled probe in the slotted section and reading the rectified crystal current. A reasonable standing wave ratio might be 2 in voltage. The position of the most convenient minimum is located with some care. Now a chart of some kind is consulted, and the impedance at the minimum is deduced to be half that of the coaxial line. A match is therefore secured by inserting a quarter wavelength of line of impedance $Z_o/\sqrt{2}$ in the line, one end of this section being at a minimum point as near as possible to the feed. There should then be a match.

We can now seek to understand this process. A very great aid in so doing is obtained from a simple geometrical construction. This is shown in Fig. 4·8. OP represents one position of the rotating vector which describes forward wave motion in the line. In the uppermost diagram this describes the voltage; in the middle diagram, the current. It is supposed that the line has an impedance unity so that the length OP is the same in both diagrams. The effect of the imperfect feed is to add a reflected wave, which

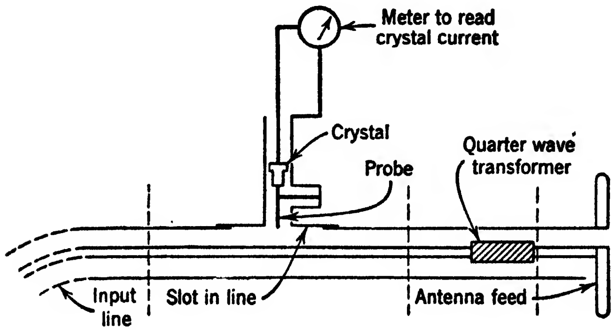


FIG. 4·7 Illustrating the process of impedance matching by measuring standing wave ratio and minimum position. The quarter wave transformer is determined in terms of these measurements and eliminates return power from the antenna.

is represented by PV' in the voltage case having a phase shift ϕ and a reflection coefficient a. As we consider points nearer the generator this phase shift increases in the manner described on page 96, taking for example the position PV'', with the increase in phase shift being $2kx$. In the current case we recall that the reflected wave has the current reversed with respect to the voltage, this being a basic property of the directions of the electric and magnetic vectors in a traveling wave. Therefore the current as seen in the middle diagram is represented by Pi' at the feed and Pi'' at a distance x down the line.

The impedance is, by definition, the ratio of OV to OI. It is real only when they are along the same line. This is only at the positions OP' for voltage and OP'' for the corresponding current value; or OP'' for voltage with OP' for current. These are the maxima and minima. In the first case the impedance is at its greatest, namely $(1 + a)/(1 - a)$, and in the second at its least,

namely $(1 - a)/(1 + a)$. The way in which the voltage and current vary along the line is shown graphically alongside each dia-

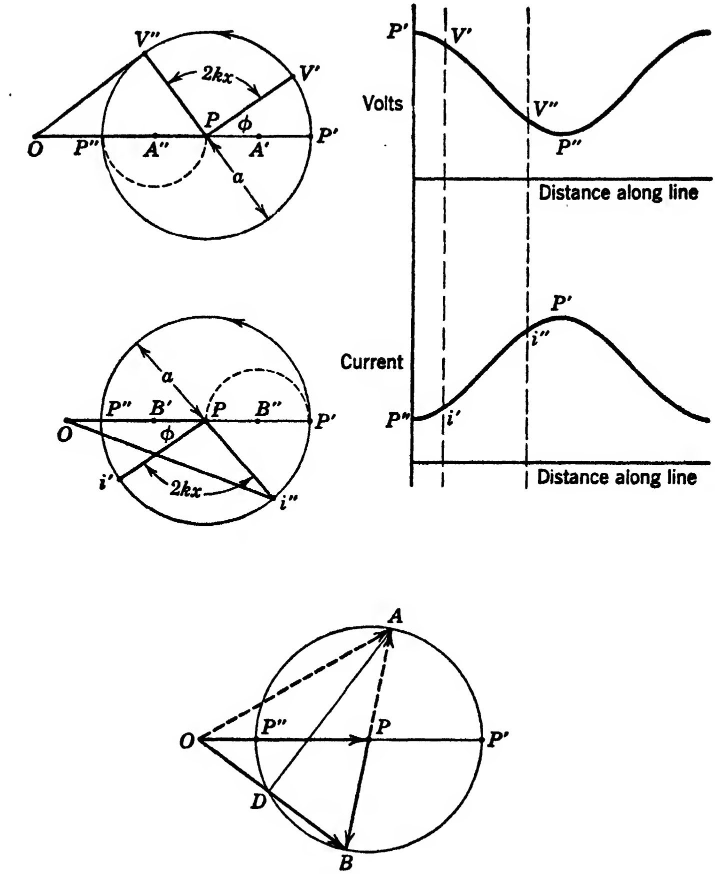


FIG. 4·8 Vector addition of transmitted and reflected waves of voltage and current. The reflected current is opposite to the reflected voltage. The vector sum of voltage is therefore great when the current is small. The ratio of these, the impedance, is real at the maxima and minima. The bottom diagram shows current (solid) and voltage (dashed) together. The impedance is the vector ratio OA/OB.

gram. It is now apparent that the standing wave ratio tells us the impedance at maximum or minimum directly. The crystal probe can be assumed to detect voltage only. Hence the maximum reads OP' and the minimum OP''. In the case we have described

the ratio of these is 2; therefore the impedance at maximum is twice that of the line and at minimum half.

To see how the quarter wave section operates we note that the new section we insert does not have unity impedance. The voltage then has to be represented by OA' and the current by OB'. These two points are now the centers of the rotating vectors for voltage and current. It can be seen that when a quarter wave has been traversed, as indicated by the dotted semicircles, the voltage and current are both at P and the impedance is at that point unity.

The Impedance Circle

The case we have considered is unusually simple. In order to make the variation of impedance in general more apparent the geometrical construction can be employed to give the basis for a valuable impedance diagram. The bottom diagram of Fig. 4·8 shows current (solid) and voltage (dashed) on one diagram. The impedance is the ratio of two vectors which are not, in general, along the same line. This fact is expressed by a complex impedance, and the two parts of the complex quantity are the real part, the ratio of OD (the component of voltage along the current direction) to OB, and AD (the component perpendicular to the current direction) to OB. The points A, D, B form a right-angled triangle, and so we have

$$BD^2 + AD^2 = AB^2 = 4a^2$$

Now we can put $BD = OB - OD$, and in addition $OD/OB = R$, the real part of the complex impedance, and $AD/OB = X$, the reactive part. With these substitutions we have

$$(1 - R)^2 + X^2 = \frac{4a^2}{OB^2} = \frac{4a^2 R}{OB \cdot OD}$$

Now $OB \cdot OD$ is equal to the square of the tangent to the circle from O which is $1 - a^2$, since OP is unity.

Rearrangement of this equation, with this substitution, then gives

$$\left(R - \frac{1 + a^2}{1 - a^2}\right)^2 + X^2 = \left(\frac{2a}{1 - a^2}\right)^2 \qquad (4\cdot 6)$$

This is a circle, radius $2a/(1 - a^2)$, with center at $(1 + a^2)/(1 - a^2)$ on the axis of R.

A similar result follows for admittance, which is derived from the ratios of the two components of current to the voltage.

These results can be made the basis for a simple impedance diagram, in which reactance is plotted versus resistance for various values of a, which is the quantity which determines the standing wave ratio. The variable on any given circle is the phase angle determined by the fraction of a half wavelength traveled down the line. Points of equal angle can be marked on the circle. The construction of these is easy because the lines of equal angle are also circles of radius $(\tan^2 kx + 1)/2 \tan kx$ whose centers are at points along the reactance axis at $(\tan^2 kx - 1)/2 \tan kx$. These circles all pass through the point of unit resistance and no reactance. This impedance diagram is useful in giving semiquantitative information and in helping to see the effect of various tuning devices.

One great advantage of such a chart is that it permits an understanding of the effect of varying the wavelength on a matching transformer. For example, consider the quarter wave transformer used previously. We know that this gives a perfect match when on frequency. Therefore we can draw a semicircle, as in Fig. 4·9, from the point $R = 0.5$, the minimum value, to the point $R = 1$. This represents the way in which impedance changes in a line of characteristic impedance $1/\sqrt{2}$ as a quarter wavelength is traversed. Now if the wavelength is increased by 10 per cent the angle traversed is 10 per cent less, and the point reached is somewhere about as indicated at C. To determine this point precisely it would be necessary to construct the lines of equal angle (i.e., to make very nearly a whole impedance chart centered on the value $1/\sqrt{2}$). However it can be seen that an estimate of the correct position can be made with fair accuracy. It is then at once apparent that a standing wave ratio of about 1.1 in voltage would be encountered. Roughly the same value will be obtained if the wavelength is diminished by 10 per cent.

We can use the chart to suggest a match which is less frequency-sensitive. Suppose, instead of matching perfectly to the point O, we match to a point beyond O at the standing wave ratio of 1.05. Then we complete the match with a second quarter wave section of the right impedance to lead us to the point O. Now if the wavelength increases by 10 per cent a glance at Fig. 4·9 shows that the first transformer is too short by 18° and the impedance

reached is at F. The second transformer then takes the impedance from there to a point which is again 18° short. This is, however, within the 1.05 standing wave ratio circle. The same is true for a 10 per cent decrease in wavelength. This illustrates one of the many possible ways of *broadbanding.*

Naturally a method which increases the bandpass by using two transformers is susceptible of still more improvement if several

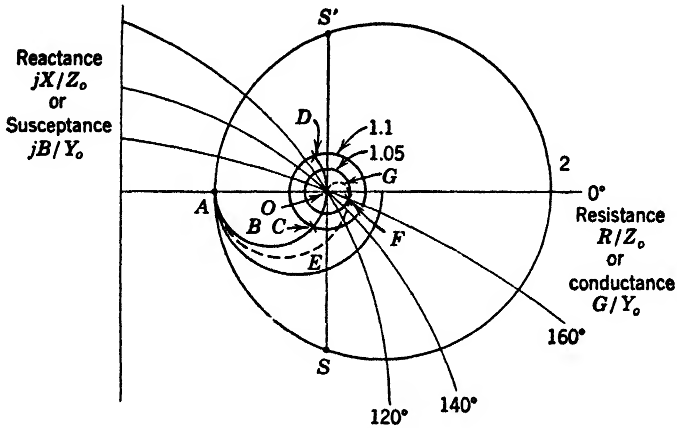


FIG. 4·9 Illustrating the use of the impedance chart. To match a standing wave ratio of 2 we can use a quarter wave transformer ABO which leaves a mismatch at C or D for a 10 per cent change in wavelength. Or two sections can be used, one matching to the 1.05 point as shown at AEF followed by another matching to O, by the transformer FGO. A 10 per cent change in wavelength still leaves the mismatch within the 1.05 circle.

transformers are used. Carried to the limit, this means that a taper section with a gentle taper is very broadband. This is found to be the case experimentally.

A very important feature of microwave technique is a length of shorted line, or stub. Its action can easily be seen from Fig. 4·10. Since the shorted end produces a reflection coefficient of -1, the point O lies actually on the circle. The impedance can thus never have any real component at all. The ends of the current and voltage vectors lie on a diameter in all cases so that the voltage vector is always perpendicular to the current vector. It can also be seen that, as the length of the line increases past a quarter wavelength, the current changes from leading the current to lagging it, always by 90°. Therefore, if we put such a stub across a

line it offers a means of varying the impedance of a line in so far as a pure reactance shunting the line is of any use. In more physical language, it is possible to take a little power out of the line and return it in the right amplitude and phase to remove the return wave.

The use of a stub can be seen from the conductance-susceptance version of the impedance chart. This is used because the stub is added in parallel with the line and for such a case admittances

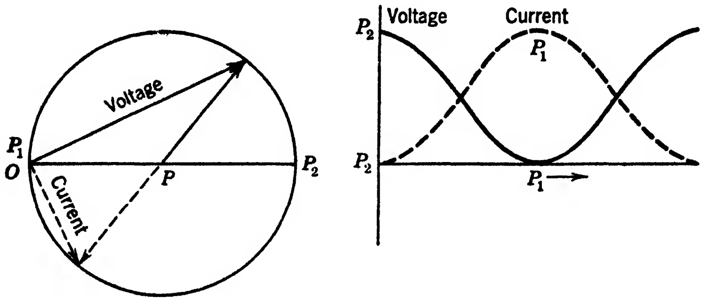


FIG. 4·10 Illustrating current and voltage relations in a shorted line. Since the voltage and current vectors are always perpendicular to one another the line has a pure reactance.

are additive. In Fig. 4·9 the same circle can be used for admittance. Then it is seen that a stub placed at the correct position on the line, as at S or S', can be adjusted to add inductive or capacitive susceptance to the line so as to reach exactly the point O and give a match.

A diaphragm placed in the line has a similar action, but adds pure reactance since it is in series with the line. The stub is more easily made variable than a diaphragm and so is used for adjustable tuning. A diaphragm is used when a permanent match is desired.

The Smith Circular Chart. This chart, which was suggested by P. H. Smith,[3] is more directly related to the simple process of vector addition and is convenient in that the value of the standing wave ratio can be included by a movable arm. The center of the diagram is the point of unit impedance ratio as before. In this case, however, the circles, in place of being circles of equal standing wave ratio and equal electrical angle, are circles of equal resistance but varying reactance on the one hand or equal react-

[3] P. H. Smith, *Electronics*, Jan. 1939, p. 29, and Jan. 1944, p. 130.

ance but varying resistance on the other hand. On the admittance chart the corresponding circles are isoconductances and isosusceptances.

It is necessary to be familiar with these charts before they become useful. It is suggested that more complete treatments be consulted if any extensive impedance matching involving different kinds of components is to be done.

Summary of Methods of Impedance Matching

In Table 4·1 is given a short summary of various methods of impedance matching. Some of these need no comment and some are discussed in the next section under the description of microwave components.

TABLE 4·1 SUMMARY OF METHODS OF MATCHING

Type of Transformer	Features
Line stretcher	Changes length of line. Does not affect standing wave ratio but may improve the power delivered.
Quarter wave section	Obeys relation $Z_T Z_0 = Z^2$. Must be inserted at a standing wave maximum or minimum. Is frequency-sensitive.
Multiple quarter wave section	Uses the same principle but can be made more broadband.
Taper section	Properly made is the limit of the above and has a wide bandpass.
Single-stub	Introduces susceptance of either sign to any amount. Can match anything if the position on the line can be chosen.
Double-stub	Can match conveniently over a wide, but not complete, range. Range can be complete with an extra quarter wave section.
Triple-stub with outer pair ganged	Can match anything.
Screw in waveguide	Adds capacitance until nearly across the guide and then rapidly becomes inductive.
Iris in waveguide	In wide side, capacitive; in narrow, inductive.
Squeeze section in guide, changing the width in the *a* dimension	Acts like line stretcher.
Termination in guide	Consists of wedge-shaped semiconductor placed across narrow dimension.
Termination in coaxial line	Poly-iron plug.

In this section we briefly describe the more commonly used component parts of a microwave system. Many different designs of these exist. What we describe are therefore largely intended as illustrations and do not constitute a complete catalogue.

Free Space

This is by far the most used and most important microwave component. All that we can usefully say about free space is, of course, included in Maxwell's equations for free space as given in Chapter 1. We can, however, inquire about the "impedance" of free space. This seems queer at first sight because no electronic currents flow. It is really a small matter because we have already made the point that, when we describe a flow of electrons, the important part is really the electric and magnetic fields resulting from the flow, particularly the magnetic field. It does not matter how this magnetic field originates; it still plays an important role in determining voltage differences, since a time-changing magnetic field produces an electric field, which is the quantity from which voltage is derived. Therefore, if we wish, we can give a more general meaning to impedance and take it to be the ratio of the electric to the magnetic vector. This property does not actually depend only on free space; it also depends on the kind of field configuration. We therefore specify that the impedance of free space is the ratio of the electric vector to the magnetic vector for plane waves of a single frequency. In the units employed in Chapter 1, statvolts per centimeter for electric field and gauss for magnetic field, there results an impedance of unity. To obtain a number in terms of ohms we require that the power transfer given by the Poynting vector $c(\mathbf{E} \times \mathbf{H})/4\pi$ is of the form E^2/Z where Z is the impedance required. We have already seen that H is numerically the same as E in the original units; therefore if we convert from statvolts per centimeter to volts per centimeter we can do so numerically for both E and H. This makes the numerical value of the Poynting vector $[E^2/(300)^2] \times (c/4\pi)$ ergs per second or $E^2/120\pi$ watts. The value of Z is thus seen to be 377 ohms. This number is occasionally, but not often, of some use.

Air

Unfortunately for many purposes free space is generally contaminated by air, which has a dielectric constant and permeability so near unity that they can nearly always be neglected. However, air can become ionized and, if electrons are once liberated and the fields are high enough, the air can pass very large currents in the form of a discharge. No simple figure can be given for the breakdown field intensity. The figure usually quoted is 30 *kilovolts per centimeter*. Everything being equal, the breakdown voltage is proportional to the pressure of air. These figures are given only for the purpose of rough calculation. It can easily be seen that in a coaxial line where the impedance is 50 ohms and the power transmitted is 10^6 watts the voltage developed is 7000 across the line. If the inner conductor has a radius of 0.5 centimeter the field near it is 40,000 volts per centimeter, which is definitely liable to cause breakdown. Therefore air breakdown is a serious matter.

Also of interest is the possibility of absorption by any constituent of the atmosphere, notably water and oxygen. This is only serious above 10,000 megacycles.

Coaxial Cable

Coaxial cable has the characteristic impedance (in ohms) of ~~130~~ log (b/a) where b is the radius of the outer conductor and a of the inner. The field developed in a coaxial line at radius r is $V/[r \ln (b/a)]$. The attenuation in such a cable is determined by the line constants: R, the series resistance per centimeter; G, the shunt conductance per centimeter; L, the inductance per centimeter; and C, the capacitance per centimeter. The attenuation, in decibels per centimeter, is then α, where

$$\alpha = 8.7\left(\frac{R}{2}\sqrt{\frac{C}{L}} + \frac{G}{2}\sqrt{\frac{L}{C}}\right) \tag{4·7}$$

Three general design considerations exist. The first is for the best distribution of the field strength to avoid breakdown. It gives a ratio of radii of 2.7 and an impedance of 60 ohms. The second is related to this and is the choice dictated by maximum power flow. Since the power flow is V^2/Z, where V is the voltage across

the line, a different ratio, namely 1.65, holds, with an impedance of 30 ohms. The third consideration is for minimum attenuation which gives a ratio of 3.6 and an impedance of 77 ohms. In practice a 50-ohm line is generally used for high power transmission and 70-ohm line for low level work.

Beads and stubs are used to support the inner conductor where solid dielectric is not used. For best transmission a line with a few stub supports is by far the best. Beads seem to have no advantage at all. Solid dielectric cable is convenient if losses of the order of 0.1 to 0.25 decibel per foot can be tolerated.

The maximum diameter of a coaxial line is determined by the possibility of non-coaxial modes (i.e., modes which are like waveguide fields). These modes can be excited at bends or stubs, and they give rise to severe attenuation and bad reflections. The greatest distance from the inner conductor to the outer should not exceed a quarter wavelength. This greatest distance is generally at a stub support.

Stub Supports

These are quarter wave sections which accordingly present very high impedance at the center conductor. Unless they are specially designed they have very narrow bandpass characteristics. The broadbanding is achieved with a stub design of the form shown in Fig. 4·11 (*a*). In broadbanding advantage is taken of the fact that, on one side of the correct length, the stub introduces an imaginary impedance of one sign while on the other side the sign is changed. Since the stub is across the line, not in series with it, it is simplest to think in terms of susceptance. This susceptance, which changes sign, is balanced against the change in length around the circle corresponding to the sleeve transformer as indicated on the susceptance diagram, Fig. 4·11 (*b*). The standing wave ratio as the wavelength is varied, is shown in Fig. 4·11 (*c*). A bandpass of 15 per cent for a ratio of 1.01 is obtained.

Such stubs can be used as straight supports or to turn right angles. The dimensions given are for ⅞-inch line.

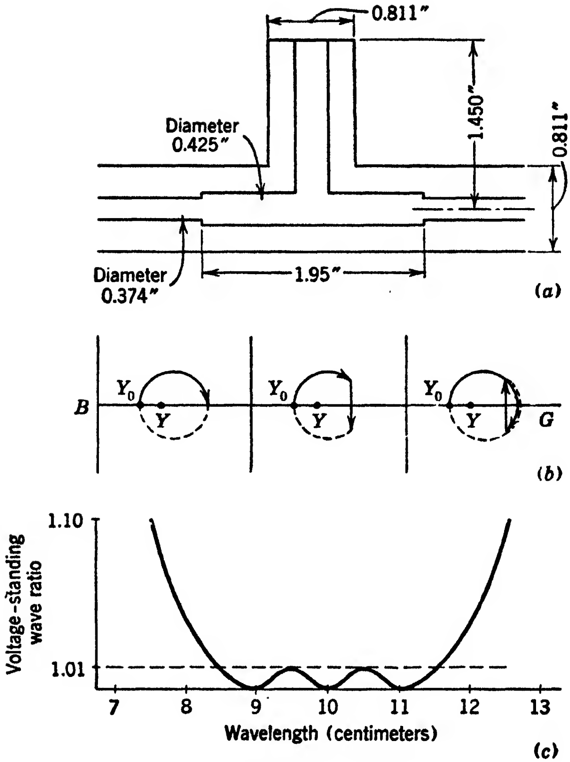


FIG. 4·11 (*a*) A broadband stub support. Its action can be followed on the susceptance-conductance chart (*b*); it is seen to be due to the fact that the stub adds susceptance of opposite signs on each side of resonance. This compensates for the change in effective length of the matching transformer as the wavelength changes. The excellent match is shown in (*c*).

Rotating Joints

These are a kind of free-for-all in quarter wave and half wave sections. A schematic design is shown in Fig. 4·12. It is found that sliding contact joints do not maintain consistent operation. Hence capacity coupling is used.

Such rotating joints can be made with a standing wave ratio of around 1.02 rising to 1.08 at 10 per cent off frequency. The breakdown conditions are good, as the regions of small spacing are

regions of low field. The critical region for voltage breakdown is the short section of abnormally small inner conductor.

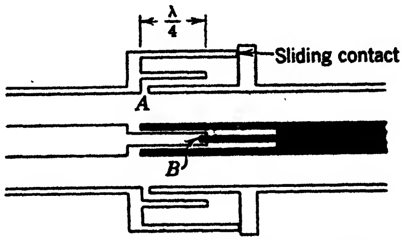


FIG. 4·12 Coaxial rotating joint. A series of 4 coaxial quarter wave chokes is used to prevent power from being radiated at the break A in the outer conductor and enable a loose contact at B, where currents are low, to connect the inner conductor.

Waveguide

Waveguide is considered only for wavelengths of 15 centimeters or less because its physical size becomes manageable only below that point. In the region from 15 to 9 centimeters there is a certain amount of option regarding the use of waveguide or coaxial line, if the power to be transmitted does not exceed 200 kilowatts. Below 9 centimeters coaxial line is troublesome because the outer diameter must be kept progressively smaller as the wavelength diminishes, in order to prevent unwanted "waveguide" modes from forming.

Waveguide does not have the property of a guaranteed single mode of excitation which is so useful in a coaxial line. For this reason the size of the guide has to be considered carefully when the wavelength to be transmitted is determined. Therefore waveguide is commonly designated as "3-centimeter waveguide," or "10-centimeter waveguide," meaning that the appropriate dimensions have been chosen for that wavelength. The basis for choice of dimensions is the selection of the critical dimension between the cutoff for the first mode and the second mode. Thus rectangular guide is chosen to have a width a [see Fig. 2·6 (d)] greater than $\lambda/2$ but less than λ. The former passes the lowest mode while the latter passes the next. Usually the width is made close to the maximum permissible without exciting the second mode. Table 4·2 gives the cutoff sizes for some modes in rectangular and circular guide.

Table 4·2 Cutoff Sizes in Rectangular and Circular Guide

Rectangular, width a depth b

$TE_{m,n}$ or $TM_{m,n}$

$$\lambda_c = \frac{2}{\sqrt{\left(\frac{m}{a}\right)^2 + \left(\frac{n}{b}\right)^2}}$$

Circular, radius a

Mode	λ_c
$TE_{0,1}$	$2\pi a/3.83$
$TE_{1,1}$	$2\pi a/1.841$
$TE_{2,1}$	$2\pi a/3.05$
$TM_{0,1}$	$2\pi a/2.405$
$TM_{1,1}$	$2\pi a/3.832$
$TM_{2,1}$	$2\pi a/5.136$

Circular guide is of interest because it offers modes of excitation which are radially symmetrical as well as modes which are

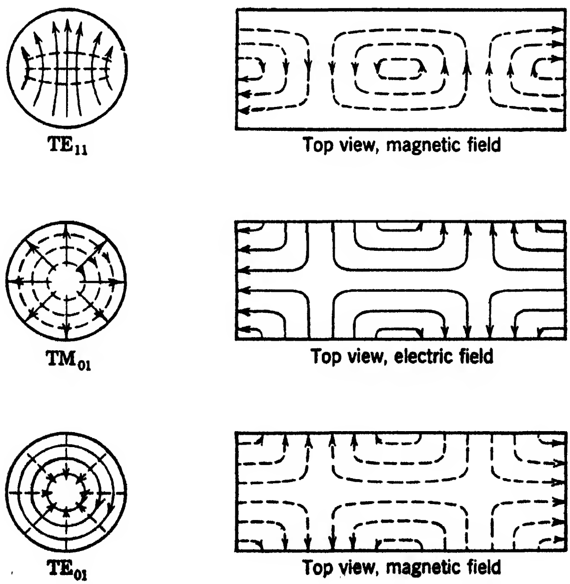


Fig. 4·13 Illustrating three modes in circular waveguide. The middle figure shows a mode which is useful for rotating joints.

very like those of rectangular waveguide. The fields for two of these are shown in Fig. 4·13, with a rectangular type given for

comparison. The TM_{01} mode is the most useful in practice because it has a relatively high cutoff wavelength, which makes it possible to excite this mode without too much complication from other modes; therefore it can be used in rotating joints.

The designation of the modes of circular waveguide is more complicated than for rectangular guide, because the same basis, namely the solution of a wave equation with the correct boundary conditions, leads to types of oscillation determined by integers. In circular guides the integers refer to the order and root of Bessel functions which are not simple to describe without elaboration.

Waveguide is very adaptable. Bends and twists are readily possible; in a bend the radius of the bend should not be much smaller than the wavelength. Usually standard bends are made up which can be soldered into the standard rectangular waveguide. The reflected power at each bend is exceedingly small, of the order of 1 per cent. A setup can therefore suffer the presence of two or three bends without noticeable reflected power. The same is true of twists.

Attenuation in guide which is well below cutoff conditions is very low. For 1½ by 3 inch waveguide it is 0.5 decibel per 100 feet at 10 centimeters, whereas for ½ by 1 inch guide it is 4 decibels per 100 feet at 3 centimeters. Insulating dirt does not greatly affect these figures as the power is still carried on the metal. However, a thin film of moisture can be most dangerous. Separate droplets are not so significant, but any pool of condensed water can give rise to an attenuation of a third of a decibel per foot or even more, in 1½ by 3 inch guide.

The *power-carrying capacity* of a waveguide is given roughly by the formula

$$\text{Power} = 3.82ab\,\frac{\lambda}{\lambda_g} \tag{4·8}$$

where the power is in megawatts, a and b are in inches, and λ and λ_g refer to the free space wavelength and the guide wavelength respectively. Then we can see that for 10.7-centimeter radiation in a 1½ by 3 inch guide where λ_g is 15.2 centimeters we have a power-carrying capacity of over 10 megawatts. It would be wise to plan on less than this as heat and processes which can put ions into the guide to start a breakdown have been ignored.

Choke Joints

It is usually impossible to design a microwave system with the entire plumbing soldered together. For example, it may be necessary to take a section out to replace a tube. Waveguide lends itself to a very simple and effective, if rather clumsy, type of joint, the *choke-flange* joint. Actually, if two pieces of rectangular waveguide are placed near each other, properly lined up, but not touching at all, the joint is remarkably good. Rapidly changing electric field intensity in the gap takes the place of conduction

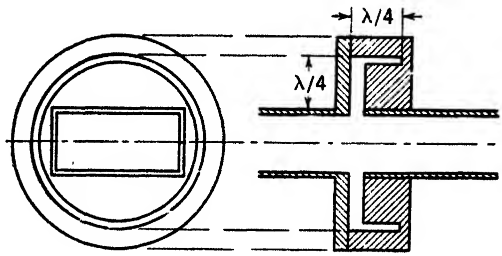


Fig. 4·14 Waveguide choke joint. When two waveguides are held end to end without contact there is nearly perfect power transfer. In the choke joint as shown this is done with the added feature of a cylindrical quarter wave choke which returns the small amount of power which has been lost.

current very efficiently. The difficulty is the mechanical problem of accurate support, and it is overcome by the choke-flange joint illustrated in Fig. 4·14. A gap is deliberately left at the junction of the two guides. The slight amount of radiation at the gap is allowed to produce a wave in a space which leads to a groove cut in the choke joint as shown. This wave is reflected at the conducting end of the groove. The total path length in this barely recognizable transmission line is arranged to be one half wavelength out and the same back, so that the reflected wave returns exactly in phase to the place where the leakage of energy occurred. It therefore contributes to the legitimate field at the gap, and the loss in the legitimate wave in the guide is eliminated. Such choke joints are so successful that losses are reduced to about 0.02 decibel per joint. They can be badly assembled or misaligned with almost no bad effects.

Tees

Tee-joints can be made in either the E plane or the H plane as illustrated in Fig. 4·15. Tees themselves are not so significant as the reason for the junction (i.e., the tee introduces a branch in the line), and how the power divides depends on what is at the end of each line. Tees are of great importance in microwave plumbing. It is worth while to point out that there is a difference between the phase relationships in the waves after division by E plane and H plane tees. These phase relationships are made use of in the "magic tee" which will be discussed later.

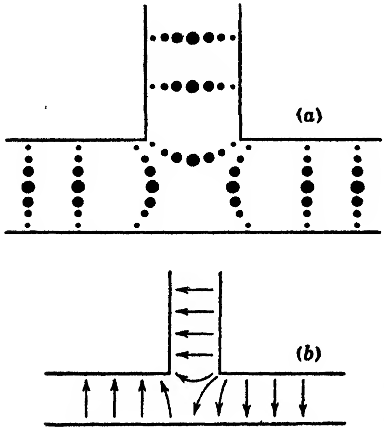


FIG. 4·15 The two forms of waveguide tees. The upper is the H plane, made in the wide dimension. The size of the dots denotes the up-and-down electric field, and it can be seen that there is no phase difference in the branches of the tee. The lower is the E plane, made in the narrow dimension. The arrows show the direction of the electric field; the phase difference in the branches is evident.

In general, some kind of impedance matching must be added to each tee. Since a tee is never included without an ulterior motive, this motive determines the nature of the match.

Feeding Waveguide from Coaxial Line

If one is interested in doing this merely for the purpose of making measurements or demonstrations, it can be achieved simply by means of an antenna placed in the waveguide and fed directly from the coaxial line. In order to achieve a good match the antenna should be placed near the end of the waveguide, which should be sealed off so that the power proceeds down the guide in one direction. The distance of the antenna from the end of the guide and the length of the antenna, or probe, are variables which have to be made optimum for a good match. This is not difficult to do and, in particular, if it is possible to put a plunger in the closed end of

the guide, the match can be made by the motion of this plunger. All this is shown in Fig. 4·16 (*a*). A broadband match can be obtained which has a voltage standing wave ratio less than 1.1 for $\pm$ 5 per cent of change in wavelength. Notice the design of the plunger which avoids contact at a high current point.

The simple method just described will handle powers up to about 100 kilowatts, but it is not very satisfactory for transmit-

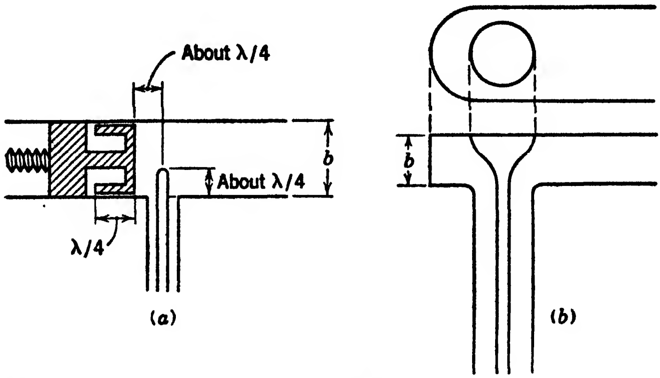


FIG. 4·16 Feeding waveguide from coaxial line. The guide is excited in version (*a*), which is adaptable for laboratory use, by a probe antenna. This is matched to optimum by a plunger as shown. For introducing high power the "doorknob" input is used. (*b*) shows a factory-matched "streamlined" input capable of handling over a megawatt. It has excellent bandpass characteristics.

ting high power. Since it is much easier to extract power from magnetron oscillators by means of the conventional coupling loop into a coaxial line, it is general practice to feed a short length of large diameter coaxial line from the magnetron and then to couple this to the waveguide. To do this a "doorknob" input is used which employs the same principle of antenna feeding, but the antenna is a kind of half wave antenna going right across the guide and making electrical contact with the upper surface. To reduce reflections back into the coaxial line, which give bad voltage problems, the antenna is spread out into a "doorknob" which is shown in Fig. 4·16 (*b*). This "streamlining" gives a very good performance as far as reducing breakdown voltages is concerned. The matching is achieved by a semicircular end plate. The original design of this is due to G. M. Hollingsworth of the General

Electric Company. The feed is as broadband as the simple design previously described.

Waveguide Rotating Joints

Two forms of waveguide rotating joints are in use. For high power systems a coaxial line rotating joint can be used which is fed from waveguide by a doorknob junction with a second doorknob junction to return to waveguide. This joint is very attractive for sets operating at above 10 centimeters since the diameter

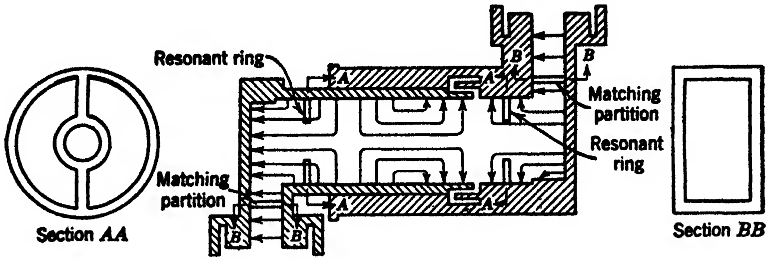


FIG. 4·17 Waveguide rotating joint. The rotation is accomplished by using the radially symmetrical TM_{01} mode. This is excited directly from the rectangular guide mode. Matching partitions reduce the reflections back into the rectangular guide, and resonant rings prevent the formation of the rectangular mode in the circular guide.

of the rotating joint is reasonable, and enormous and expensive bearings are not needed.

A second form which is in considerable use is shown in Fig. 4·17. A transition is effected to the TM_{01} mode in cylindrical guide as already described. This mode has radial symmetry and so permits rotation without alteration of the fields. The ordinary $TE_{1,0}$ mode enters through a choke joint, and is matched by a diaphragm at *BB*. The field due to this spreads into the cylindrical guide and excites the TM_{01} mode. The opposite occurs at the other end where the TE_{10} mode is re-excited in the output guide. Since the cylindrical guide is large enough to permit the cylindrical version of the rectangular guide mode, which is not radially symmetrical, it is advisable to cut this mode out by means of rings which will permit the axial mode but not the usual transverse mode. About 7 per cent improvement in match is

obtained in this way. Such rotating joints are effective and reliable and are in general use for wavelengths below 10 centimeters.

Diaphragms in Waveguides

It is a relatively easy matter to insert diaphragms in waveguides. They can be shaped to enable the appropriate impedance match to be attained. Such diaphragms therefore play an important part in microwave technique.

A large number of diaphragms are capable of theoretical calculation. The results are generally expressed in terms of an "equivalent circuit." Such calculations were made by the theory section of the Radiation Laboratory and appeared in the *Waveguide Handbook*.[4] Other information is contained in the Sperry *Design Data* book already quoted. For actual design purposes such sources should be consulted. Here we give only very broad information on some interesting types of diaphragm with approximate formulas which might enable rough tuning to be accomplished.

The first two diaphragms are the capacitive and inductive types shown in Figs. 4·18 (*a*) and 4·18 (*b*). Their equivalent circuits are also shown. The value of B, the shunt *susceptance* for the capacitive case, is given approximately by

$$\frac{B}{Y_0} = \frac{9.2b}{\lambda_g}\left(\log_{10} \csc \frac{\pi\delta}{2b}\right) \tag{4·9}$$

where λ_g is the guide wavelength, Y_0 is the admittance of the unobstructed guide, and δ and b are as shown.

For the inductive case the *reactance* X is given by

$$\frac{X}{Z_0} = \frac{a}{\lambda_g} \tan^2 \frac{\pi\delta}{2a} \left\{1 + \frac{3}{4} \sin^2 \frac{\pi\delta}{a} \left[\frac{1}{\sqrt{1 - \left(\dfrac{2a}{3\lambda}\right)^2}} - 1\right]\right\} \tag{4·10}$$

where Z_0 is the impedance of the unobstructed guide and δ and a are as shown. The case of a tuning screw illustrated in Fig. 4·18 (*c*) is of interest as it provides a very convenient method of

[4] Radiation Laboratory Report 43, Feb. 7, 1944.

tuning an experimental waveguide arrangement. The screw adds susceptance slowly at first, then rapidly until a point about 0.8 times the height of the guide is reached. It then changes sign and becomes inductive. This would be expected because a vertical obstacle which crosses the entire guide is inductive.

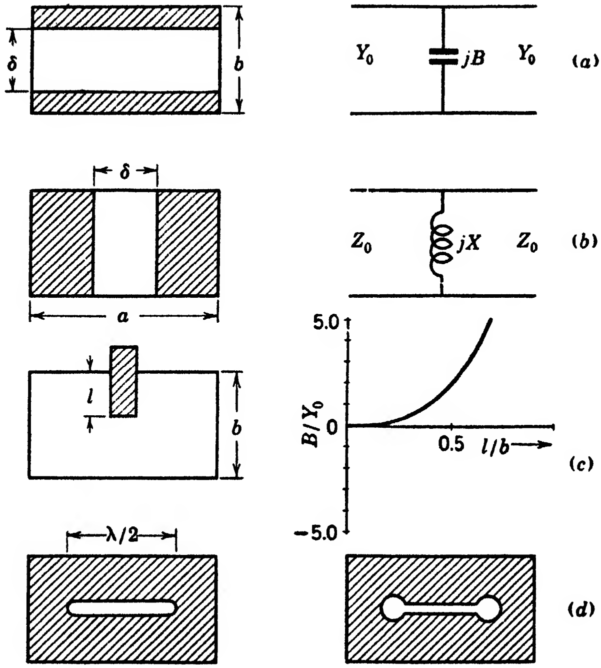


FIG. 4·18 Some waveguide diaphragms: (*a*) is capacitive; (*b*) is inductive; (*c*) is a schematic drawing of a tuning screw which largely adds susceptance as indicated; (*d*) shows two kinds of resonant slit.

The general nature of the reactance change is shown in Fig. 4·18 (*c*).

Resonant slits are of considerable interest. Two varieties of resonant slit are shown in Fig. 4·18 (*d*). These are slits which transmit at a particular wavelength. For the straight slit shown, the wavelength is very slightly longer than twice the slit length. The frequency selectivity Q of such slits is generally in the neighborhood of 10, although a very narrow slit can have a Q of 50. Resonant slits find application in gas discharge switches where the

narrow portion of the slit can be used to develop a high field which causes the gas to break down and completely alter the field distribution as a result of the current flow across the gap. A dumbbell type of slit is suitable for this purpose.

Waveguide Impedance

In the above discussion of the effects of diaphragms, the impedance of the waveguide appeared for comparison with the reactance introduced by the diaphragm. The impedance of the waveguide, like that of free space, is not a definite quantity due to the possibility of various modes. Actually it is not often of importance to know the explicit value of the waveguide impedance, since the process of matching is a relative matter. Using the same definition of impedance as that produced for free space, namely the ratio of the electric field in volts per centimeter to the magnetic field re-expressed in terms of current equivalent, we have, for the impedance of a guide carrying a TE mode, $Z_0 = 377\ \lambda_g/\lambda$ ohms, and, for a TM mode, $Z_0 = 377\ \lambda/\lambda_g$ ohms. If the medium in the guide has dielectric constant K and permeability μ, these values are multiplied by $\sqrt{\mu/K}$.

Tuning a Waveguide System

For experimental arrangements tuning screws are very convenient, and they serve the purpose excellently. Two screws placed ⅜ guide wavelength apart will match almost any line. For high power these are not satisfactory, as they are liable to cause breakdown. Some use has been made of double-stub tuners, designed after the pattern of coaxial line tuners. These are short sections of waveguide joined as E plane tees and spaced $\frac{3}{8}\lambda_g$ apart. Such tuners have been known to survive half a megawatt at 10.7 centimeters wavelength. If the plunger is not carefully made they are liable to break down. The best procedure for high power applications is to survey the line at low power and match it correctly. It should then be satisfactory at high power. If this expectation is not realized, the best practice is to find out why, rather than to attempt to construct a high power impedance matcher.

A very important phase of microwave technique is concerned with radiating microwave energy into space. This is paramount in radar and radio communications. It is quite important in laboratory experimentation. For example, the original observation of the absorption of microwaves by ammonia was carried out in the space between two reflecting mirrors. The development of the art of creating beams of microwaves is therefore of prime interest.

Microwaves, although they are, of course, of the same fundamental character as light waves, differ considerably regarding the basis for design of reflectors. The difference is simply that between geometrical optics and physical optics. In the case of visible light it is customary to point out that light consists of waves, and that wavefronts are curved by refraction in lenses or by reflection at mirrors; but the actual calculations of lens and mirror behavior are almost always made with formulas derived from geometrical optics. The reverse emphasis is found in microwave optics. One may quote a focal length to a paraboloid, and occasionally use reasoning which is founded on geometrical optics; but the fact that the wavelength of the radiation is not far from the dimensions of the reflector forces the consistent use of physical optics. Thus it is clear that the phenomenon of reflection from a paraboloid is one of diffraction in the main, and not of geometrical reflection. We thus expect to find the relations of physical optics a great aid in securing narrow microwave beams.

So many microwave antennas involve illuminating a paraboloid to obtain high directivity that there is no doubt that first consideration should go to considering paraboloid radiation patterns.

Patterns from Paraboloids

If a reflecting surface is illuminated by a source of small dimensions, there is a distribution of energy in the reflected beam. This consists of a maximum, along the direction expected according to simple ideas, with subsidiary maxima which are determined by the size of the reflector, the wavelength of the radiation, the way in which the illuminating energy is distributed, and the angle of

incidence of the radiation on the reflector. This pattern of maxima resulting from such illumination is called the secondary pattern. The way in which the illuminating energy is distributed is called the primary pattern.

The secondary pattern is due to the summed effect of the illumination over the surface; or in other words, to the primary pattern as defined by the reflector. This summation includes amplitude and phase. In general amplitude depends on the distance traveled, and phase on the distance divided by the wavelength. A source placed at the focus of a paraboloid will therefore produce disturbances after reflection which are equal in amplitude and phase. This is a long way of saying that a plane wave is developed which does not diverge, a result which is obviously desirable. The extensive use of paraboloids is due to this fact. It may be thought that there is no problem other than locating the focus and employing as near to a point source as possible. This is almost true in the optical analogy, but with microwaves, where the wavelength compares with the reflector, this step is only a beginning. We have already said that there exist maxima and minima due to diffraction. These maxima are called *sidelobes.* Such sidelobes may ruin discrimination in a radar, or cause "crosstalk" in radio communications. Therefore it is also necessary to pay attention to means of illuminating a paraboloid so as to eliminate or reduce sidelobes. While doing this it is naturally necessary to retain high directivity. There is therefore considerable art to the illumination of a paraboloid. We give here a few of the simple considerations of that art.

We first tabulate some definitions:

Antenna gain is the ratio of the peak power of the reflected beam to the average (i.e., to the power obtained if it were uniformly distributed over a sphere).

Antenna gain function is the ratio of the power radiated in the solid angle defined by the azimuth angles θ and $\theta + d\theta$, and the elevation angles ϕ and $\phi + d\phi$, to that of an isotropic radiator between the same angular increments. Explicitly this is

$$G(\theta, \phi) = \frac{\text{power per unit solid angle at } \theta \text{ and } \phi}{\dfrac{\text{total power}}{4\pi}}$$

E plane and *H plane* are defined with respect to the polarization of the feed; they denote the planes of the electric vector and the magnetic vector of the feed respectively. For a feed which is the equivalent of an oscillating dipole the E plane contains the dipole axis and the H plane passes through the center of the dipole perpendicularly to it.

Half power width is the full angular width of the central maximum of the power antenna pattern (power as a function of angle), either transmitted or received *one way*, measured at the angles for which the power is one half peak power.

With these definitions to aid our language we can proceed.

Factors Determining Gain and Sidelobes

The highest gain that can be obtained from a paraboloid is that resulting from uniform illumination of its surface. This, however, also gives high sidelobes. No sidelobes at all are obtained for an illumination which is a Gauss error function of the distance from the center of the paraboloid. Such illumination is only possible approximately and then is very wasteful of the area of the paraboloid so that the gain is very low. Somewhere in between is an optimum. A most convenient way of considering the effect of the distribution of the energy from the feed is due to Spencer,[5] who considers illumination functions of the form

$$E_P = I_0 \left[1 - \left(\frac{r}{R}\right)^2\right]^p \tag{4·11}$$

where r and R have the meanings described in Fig. 4·19, which is intended to make clear the meaning of the term illumination function. The greater the value of p the more rapidly does the illumination diminish toward the edge of the paraboloid. Uniform illumination is obtained for $p = 0$. In Fig. 4·19 the plane ACB, taken reasonably close to the paraboloid, is a plane of equal phase, if the source is effectively a point and is at the focus. The illumination over ACB is the primary pattern, from which the secondary pattern at large distances is developed. The secondary pattern is described in terms of the angle θ as marked.

[5] R. C. Spencer, Radiation Laboratory Report T-7, Oct. 1942.

The distribution of energy along a radius r from C will depend on the nature of the feed. Its distribution is represented by shading and by a bell-shaped function. This bell-shaped function is described by equation 4·11.

Using this function and applying the fact that the secondary pattern, which is described in terms of θ, can be derived from the

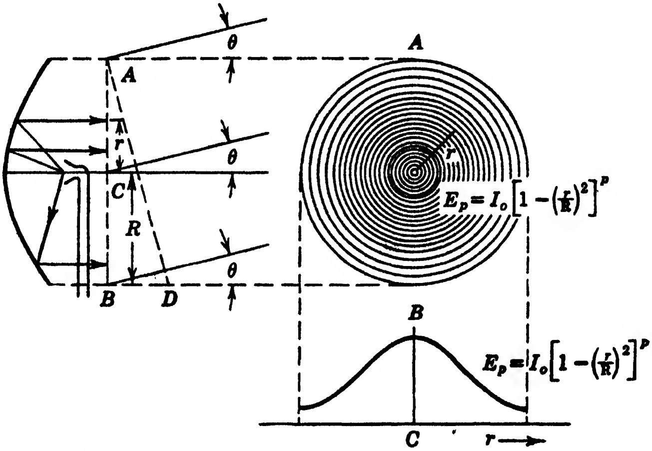


FIG. 4·19 Geometrical picture of a paraboloid and feed. The property of the focus of a paraboloid guarantees that along line AB the phase is constant. The primary pattern of the feed gives an amplitude distribution over the plane of the aperture which is indicated by the shading of the circle and by the contour below. The secondary antenna pattern is due in part to the nature of the primary pattern and in part to its interruption by the aperture.

primary pattern by means of the Fourier integral,[6] Spencer has prepared the table shown here as Table 4·3.

TABLE 4·3

p	Gain	Half Power Width (W)	First Minimum	First Subsidiary Maximum	Intensity of First Maximum Relative to Main Maximum
0	$9200D^2/\lambda^2$	$1.87\lambda/D$	$2.28\lambda/D$	$3.06\lambda/D$	1.75%
1	$6900D^2/\lambda^2$	$2.31\lambda/D$	$3.06\lambda/D$	$3.79\lambda/D$	0.34%
2	$5100D^2/\lambda^2$	$2.68\lambda/D$	$3.79\lambda/D$	$4.52\lambda/D$	0.09%

In this table the angular values are in degrees, the diameter of the paraboloid, D, is in *feet*, and λ, the wavelength, is in centimeters.

[6] See Appendix 1.

In addition it is useful to remember that the gain of a uniformly illuminated paraboloid, for which $p = 0$, is

$$G = \frac{4\pi A}{\lambda^2} \tag{4·12}$$

where A is the area of the paraboloid, in the same units as λ^2. Also the product of gain and the square of the half power width, GW^2, is nearly a constant. If W is expressed in radians the constant is unity. If further sidelobes are of interest, it can be shown that they decrease as $[\lambda/(2\pi \sin \theta)]^{2p+3}$ in power.

Table 4·3 is very informative. In the first place it is apparent that very high gains can be obtained. By making the paraboloid of diameter 10 feet at 10 centimeters wavelength the gain for $p = 1$ is 6900. About two thirds of this can actually be realized. Illuminating such a paraboloid at 1 centimeter gives a gain one hundred times larger. The highest antenna gain in use to the authors' knowledge is in a lightweight height finder radar which has a gain of 18,000 and employs a 10 by 3 foot elliptical section of a paraboloid to give a spread beam in one direction. The radar operates on a wavelength of 3 centimeters.

In the second place the sidelobes can be reduced to very low figures if the illumination is tapered off toward the edge in the optimum way. Thus a first sidelobe of intensity 30 decibels below the peak intensity is quite realizable without too much sacrifice of gain.

If a paraboloid cut to a non-circular form is used, as in the case just cited, the figures given in Table 4·3 become approximate only. They are still of some use if the gain is expressed in terms of area, as in equation 4·12, rather than in terms of diameter. Moreover, the widths of the beams in the two dimensions can be estimated qualitatively by treating each dimension as though it were a true diameter.

Actual Antenna Designs

The actual antenna pattern from a 30-inch-diameter dish illuminated at 3.2 centimeters using a waveguide feed with a disk reflector is shown in Fig. 4·20. This figure is taken from an article by G. F. Hull, Jr.[7] It can be seen that it agrees in broad char-

[7] G. F. Hull, Jr., *Am. J. Phys.*, 15, 111 (1947).

acteristics with the figures of Table 4·3. Such patterns are experimentally plotted by using the antenna as a receiver and taking advantage of the *reciprocity theorem.* This theorem can be thought of as a consequence of Huygens' principle, and it states that interchanging the transmitter and receiver does not change the ratio of the power transmitted to power received. In terms of antenna patterns this means that the pattern for reception is

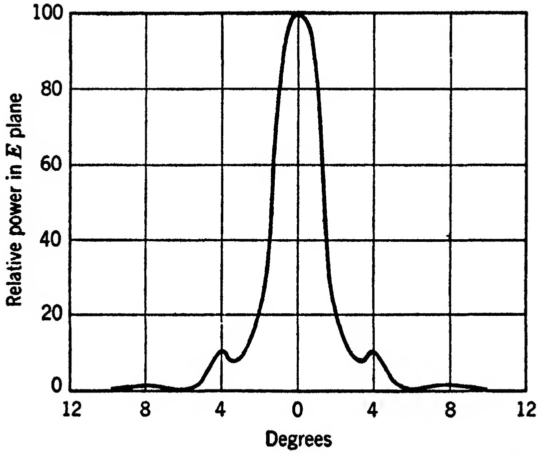


FIG. 4·20 Experimental power distribution pattern from a 30-in. paraboloid illuminated by 3.2 cm radiation.

the same as for transmission, and so it is possible to measure the pattern by observing the power received by the antenna from a transmitter located at a considerable distance. The antenna is rotated into a series of directions, and the pattern is plotted.

Feeds

As stressed previously, the actual antenna pattern is determined largely by the primary pattern, which depends on the nature of the feed, and many designs have been tried. Some of these are shown in Fig. 4·21. Almost any desired primary pattern can be obtained, particularly for feeding a paraboloid.

A *horn feed* illumination is the most widely favored. A horn is simply a splayed end to a waveguide. Horn feeds are reasonably easy to match and have moderate bandwidth. An interesting

form of feed for ground radar is the linear array, which is described later.

The power-handling capacity of a feed is of some interest. A megawatt may have to be dealt with. The energy flow in watts per square centimeter is $E^2/377$, where E is in volts per centimeter. We can calculate the minimum area across which the megawatt must flow, assuming that breakdown occurs when E is 30,000. It

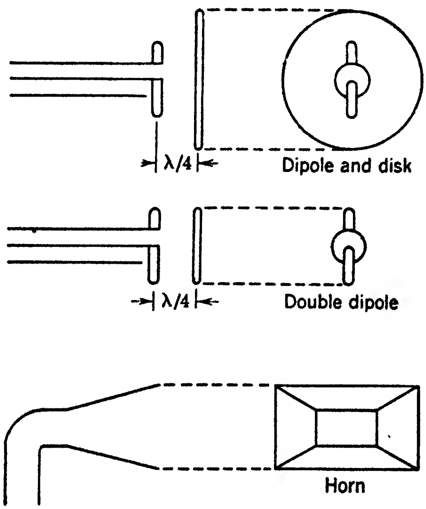


FIG. 4·21 Three common types of microwave antenna feed.

is seen to be about ½ square centimeter. In view of the fact that fields near a dipole are not uniform and that some safety factor has to be allowed in case there is a mismatch, it can be seen that breakdown can easily occur. Hence it has proved absolutely necessary to keep pressure in all high power microwave feeds unless they are well distributed, as in the linear array. A horn lends itself very readily to this as it may be closed by a single thin polystyrene plate which does not require elaborate matching.

The paraboloid itself must be rigid and geometrically right. In large paraboloids this is a limiting factor. The doctrine usually held is that the "dish" should be accurate to within 1/16 wavelength.

One point of importance in feed design is the interference effect of any back lobe in the feed. This is in the direction of the main

beam, but it has not been reflected from the paraboloid and so may not be in the right phase relationship with the rest of the energy. Abnormal sidelobes may result. To reduce these, the focal length of the paraboloid should be a whole number of half wavelengths, which restores the right phase relationship.

This same effect of interference between the reflected energy and direct energy is responsible for much of the trouble of matching an antenna over a wide band of frequencies. Power is inevitably reflected back into the line if the feed is on the axis of the paraboloid since reflected energy must return to it. By feeding from off the axis and tilting the dish so that no energy is returned to the feed, as can be done in some applications, the problem of matching is greatly reduced.

Beam Shaping

The simplest method of shaping the beam by a small amount is to use the radiation from two paraboloids. This is done by tacking a strip of metal, shaped to a different focal length, in the appropriate position to radiate the power as wanted. This may, for example, appear as a "chin" on a paraboloid to give more upward radiation, or a vertical strip to give a more fanned beam. If the shaping required is considerable the process of obtaining the right pattern is very laborious.

Usually the fanning is required merely to give an increased width to the beam; for example, in shipboard radars with unstabilized antennas where the roll of the ship makes detection intermittent unless the beam is fanned considerably. The considerations of Table 4·3 show that for such fanning the "aperture ratio" must be high (that is, the dish must be longer in one direction than the other). To illuminate such a dish is difficult. Highly frequency-sensitive methods can be devised; for example manufacturing the dish in the form of a "cheese," which is a cylindrical paraboloid with sides to it. When horn-fed this gives aperture ratios as high as 5 to 1, but has poor bandwidth. The next advance was made by using a linear array to feed a cylindrical paraboloid. The aperture ratio available by this means is well over five. Such an antenna has one serious disadvantage in that the direction of the beam is dependent on frequency. Often this condition is not troublesome and, where it is not, this method is excellent.

It has turned out that a paraboloid cut to nearly the shape of the primary pattern of a horn feed can be made with an aperture ratio of three or more. Such antennas have quite high gain, good bandwidth, and are often quite satisfactory.

Illumination off axis is very liable to increase the sidelobes. The beam deterioration is so great that in general it can be said that distorting the beam by as much as one beam width by off-axis illumination is the maximum allowable.

Linear Arrays

If a series of dipoles is placed at regular intervals along a waveguide and coupled to the field in the guide by short probes, the phase of the radiation from each dipole is determined by the position on the waveguide and the wavelength of the radiation in the guide. The power radiated by a series of such dipoles is given by an expression virtually identical with that for the light intensity at a given angle to a diffraction grating. The important term in this is

$$\text{Intensity} = A\,\frac{\sin^2 Nnd}{\sin^2 nd} \tag{4·13}$$

where N is the number of dipoles, d is the distance between them, and n is a quantity defined by

$$n = \pi\left(\frac{1}{\lambda_g} - \frac{1}{d} - \frac{\sin\theta}{\lambda}\right) \tag{4·14}$$

or

$$n = \pi\left(\frac{1}{\lambda_g} - \frac{1}{2d} - \frac{\sin\theta}{\lambda}\right) \tag{4·15}$$

according to whether the dipoles are always fed identically or are reversed alternately. θ is the angle made with the normal to the guide, λ and λ_g are the free space and guide wavelengths respectively.

The radiation maxima are found from equation 4·13 to be determined by the vanishing of the denominator, which leads to the requirement that either

$$\sin\theta_{\max} = \frac{\lambda}{\lambda_g} - \frac{\lambda}{d}(h+1) \tag{4·16}$$

or

$$\sin \theta_{\max} = \frac{\lambda}{\lambda_g} - \frac{\lambda}{d}\left(h + \frac{1}{2}\right) \tag{4·17}$$

where h is 0, 1, 2, etc.

These are two very interesting expressions. The guide wavelength is always greater than the wavelength in air, and so the only possible solutions of 4·16 are those which give maximum radiation along the guide, the so-called "*endfire array.*" This can

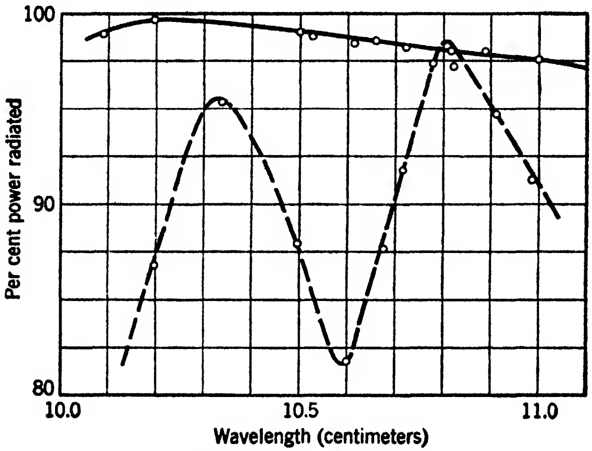


FIG. 4·22 Power radiated as a function of wavelength fed in, for a 105-element broadside array feeding a cylindrical paraboloid. The full line is for off-axis feed so that no power is fed back into the feed from the "dish." The dashed line is for a similar feed which is on-axis and so is subject to return power. The effect on the bandpass is striking.

be used for some broad beam antennas where high gain is not needed. If the dipoles are reversed alternately, formula 4·17 applies, and with the right choice of d it is possible to make $\sin \theta_{\max} = 0$ and so obtain a *broadside* maximum. It turns out that the exact broadside gives too high a standing wave ratio to be easily matched, so that a small angle of "squint," of the order of 15° to 20° has to be tolerated. Such an angle gives a feed of low standing wave ratio with exceptionally good bandpass. The bandpass of a 105 dipole array, feeding a cylindrical paraboloid along the axis of the cylinder but off the paraboloidal axis, is shown in Fig. 4·22 as the solid line. For comparison the on-axis power curve is

shown. The bad effect of the return power entering the feed is striking. The data are Blackmer's.[8]

Rapid Scanning

The linear array has an important application suggested by Alvarez and developed further by Clapp, among others, namely that by changing the wavelength in the guide the angle of the maximum can be varied and the beam will be scanned. The change in wavelength can be made by varying the primary frequency of the transmitting tube. This method of scanning may prove to be very excellent when tunable magnetrons are readily available with convenient methods of tuning. The other method is to vary the width of the guide, the so-called "delta-a" process. An increase of 50 per cent in width of the guide makes possible an angle of scan of 20°. This is done in the GCA, which is described in Chapter 11, and in other specialized radars. Other methods of rapid scanning are available, but are somewhat elaborate to describe.[9]

4·5 POWER MEASUREMENTS

The measurement of power is important in microwave technique. It is only by careful measurements of power that the actual behavior of a microwave "circuit" can be accurately described.

High level and low level measurements differ markedly. The difference lies in the fact that at low level we are interested in amounts of power which are at the limit of detectability and so have to use maximum amplification. At high level there is no problem of detection, and the main considerations are accuracy, reliability, and convenience. We can consider high level first.

Water Load Measurements

The fact that water readily absorbs microwaves can be made use of for bench calorimetric determinations of microwave power.

[8] L. L. Blackmer, Radiation Laboratory Report M-156-D, 1944. For general information on linear arrays see H. J. Riblet, "Microwave Omnidirectional Antennas," *Proc. I.R.E.*, **35**, 462 (1947).

[9] Further information on antennas can be found in the Radiation Laboratory Technical Series and in "Radar Antennas" by H. T. Friis and W. D. Lewis, *Proc. I.R.E.*, **35**, 219 (1947).

The microwave power is converted into heat, which is measured by measuring the flow of water and the temperature rise due to the absorption of power. The heat produced in calories per second is so measured. It can readily be converted into familiar wattage units. The only problem is therefore that of matching into the water load. In one form of design the water load is a section of coaxial line with a glass window one quarter wavelength long sealing the main line from the water section. The glass is chosen so that the quarter wave thickness acts as a matching transformer. In a waveguide a simple method of matching is to place a glass section containing the water diagonally across the short side of the guide. This is a familiar method to secure attenuation without reflection and is commonly applied as a form of termination where power must be dissipated and no power returned to the line. Such water loads can be used to measure average powers of 10 watts and up. They are accurate, and they are the ultimate reference for calibration.

Directional Couplers

A directional coupler is a device which abstracts a definite small fraction of the power flowing in one direction in a line but does not respond appreciably to power flowing in the other direction. In the early days of microwave development the simplicity and scope of standing wave measurements impressed all workers. As the subject advanced into maturity ingenuity began to be applied to the problem of actually isolating a wave in one direction from the other. Ultimately so many designs of "directional coupler" appeared that a whole report was necessary to index them. Granted that such a device exists, it can be used to monitor power continuously by taking a known fraction of the total power and feeding it into a thermistor connected to a bridge. The fraction of the power fed to the thermistor must be determined by calibration with a water load. After this has been done the directional coupler and thermistor form a satisfactory means of monitoring power. For a discussion of this see a report by Hadley.[10] A brief account of directional couplers is given in the next section.

[10] C. F. Hadley, Radio Research Laboratory Report 411-246, Sept. 1945.

Low Level Power Measurement

The instrument which makes low level power measurement possible, if one has sensitive receivers, is an accurate attenuator. If power of the order of micromicrowatts is to be measured, there is first the problem of its detection. We can suppose this to be achieved by some sort of low noise amplifier. It is then possible to use this amplifier to observe the signal resulting from the attenuation of a known amount of power fed in from a relatively high power source through an attenuator. By comparison of the signal to be measured and a signal of equal amplitude from the synthetic source through the attenuator it is possible to make quite accurate measurements of power. Thus for the measurement of the echo strength observed in a radar, the echo itself is matched with a synthetic echo from a signal generator which is fed through the attenuator and the attenuator is read. Measurements of this kind can be made fairly easily, for example while on an airplane in flight, and can be used to determine the performance of a radar.

4·6 OTHER MICROWAVE DEVICES

In this section we describe briefly some microwave equipment which was developed primarily as part of radar research and which seems likely to have future application in radar or elsewhere. The items concerned are cavity gas switches, or "TR boxes"; r-f amplifiers; directional couplers; "magic tees"; and "rat races."

Cavity Gas Switches or TR Boxes [11]

To keep the high power out of the receiver of a radar while the outgoing pulse is in the r-f line, a cavity gas switch was devised by Sutton at Oxford University and by Lawson at the Radiation Laboratory almost simultaneously. The purpose is to feed the high power radiation into a cavity resonator of high Q where the voltage developed is enough to cause a discharge in the gas in the

[11] As far as the authors can tell the word "box" reverts to the days when the TR was not yet invented and was referred to as a black box. TR of course is the abbreviation of transmit-receive.

cavity. This causes a vigorous redistribution of current which may be made to act as a switch.

The general appearance of such a switch is shown in Fig. 4·23. It consists of a tube containing gas which can be introduced into a cavity. The tube itself contains hydrogen and a little water vapor at about 1 centimeter of mercury pressure. The two electrodes are sealed into the glass envelope, leaving two flanges to which the cavity can be attached. The electrodes are hollow cones

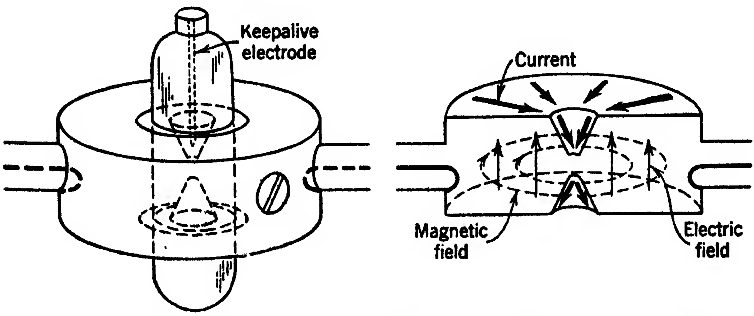


Fig. 4·23 Cavity gas switch or TR tube. The production of a high field between the conical electrodes starts a discharge which changes the effective manner of operation of the cavity to that of a coaxial type with a radial electric field. This completely upsets the power transfer conditions and can be made to protect a sensitive receiver.

with about a millimeter separation at the gap. The starting of the discharge is facilitated by a "keep alive" electrode which maintains a steady low current discharge inside one of the cones, thus providing a permanent supply of ions in the discharge space. The cavity can be tuned by tuning screws of about ⅝ inch diameter. The tube described is the 721. The 1B27 is similar but tuning is done by changing the gap spacing by means of a metal tuning screw sealed into the envelope.

The manner of action is as follows. When no discharge passes, the cavity is excited by the input coupling loop as shown, with the electric field along the axis of the cylinder having a high value at the gap. If a discharge passes between the ends of the cones the conditions are profoundly altered. Considerable conduction current is now flowing through the gas across the gap; this current will modify the electric fields in that region so that the boundary conditions which determine the manner of oscillation of the

cavity are no longer obeyed. To cite an extreme version of the process we can suppose that the effect of the discharge is to replace the gap by a conductor. If this is done the electric field must be *radial,* and not as drawn; moreover, the resonant wavelength is twice the height of the cavity, which is far from the operating frequency. There is therefore a strong mismatch to both lines. The cavity in the discharging state is thus a means of rejecting the flow of radiofrequency, whereas in the non-discharging state it is well matched and causes very little reduction in the power delivered.

TR tubes can reduce 200 kilowatts in the main line to 50 milliwatts on the far side, while permitting a loss of only 2 decibels when a discharge is not present. They have a "clean-up time" of up to 200 microseconds if measured at the highest power and the greatest receiver sensitivity. This is due to the existence of residual ions in the discharge after the pulse has passed. TR tubes have a life, under exacting conditions, of about 200 hours. Since the process we have described does not take place instantaneously, a TR tube passes a short but intense "spike" at the start of a pulse which is troublesome.

Bandpass TR

A recent design of TR by M. D. Fiske at the General Electric Company is likely to displace the above type of TR. Discharges are produced in resonant slits, a device originally used by Longacre in an early high power ground radar. If a resonant slit such as Fig. 4·18 (*d*) is placed in a waveguide the high power will cause breakdown, and there will be ion current in the gap which is then effectively closed. Without the discharge the transmission is virtually complete for the correct frequency.

Such air gaps were tried by Longacre with some success, except for short lifetimes on account of burning at the gap. A design for an enclosed, partially evacuated gap was then worked out, and this was given the name "beetle."

A considerable improvement on this design was made by Fiske, who showed that three such beetles placed at a quarter guide wavelength spacing gave very broad bandwidth and low leakage power under discharge, in addition to having a low transmission loss.

R-F Amplifiers

One of the severest limitations in the whole of microwave technique is the great difficulty of amplifying microwave power. Amplifiers do in fact exist, but either they are limited as to power or they are noisy.

A GL446 lighthouse tube can be used as an r-f amplifier; one set up for such use is shown in Fig. 4·24. The noise figure in such an amplifier increases rapidly as the wavelength diminishes below 20 centimeters, and the gain decreases correspondingly. There is

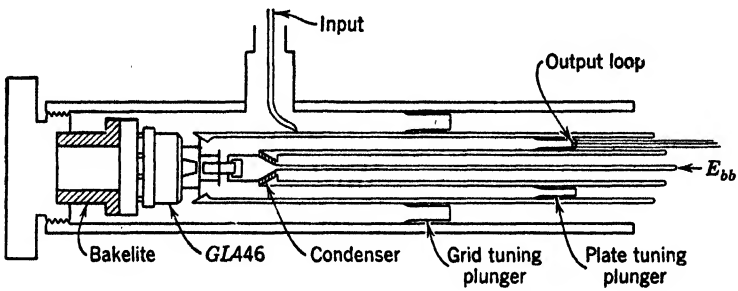


FIG. 4·24 Lighthouse tube as r-f amplifier. Fully separate grid and plate cavities are used. Input is by contact to the inner conductor of the grid cavity. Output is from a loop near the tuning plunger in the plate cavity. This is only one of several possible arrangements.

no reason to believe that this limitation is inherently insurmountable. An experimental low power amplifier has, in fact, been built by Neher. A power gain of 15 to 30 decibels is obtained at a bandwidth of 6 megacycles and a noise figure (page 246) of 10 to 14 decibels below theoretical.

Directional Couplers

If a line has waves running in two directions it is possible to couple a second line to it in such a way that there is asymmetry between the two directions and the added line. By a suitable design the effect of this asymmetry can be made such that only one of the running waves is found in the second line. The original name for this device was "wave selector"; at present it is called directional coupler.

Three such directional couplers are shown in Fig. 4·25. The information for this drawing is taken from a report by Severinghaus.[12] The first is the "Bethe hole" type of coupler. In it a second waveguide is placed across the main guide at an angle deter-

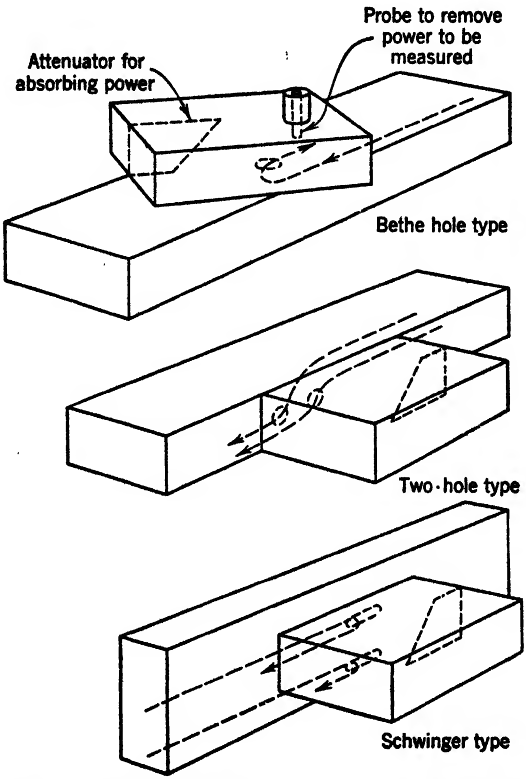


FIG. 4·25 Three types of directional coupler. The arrows indicate the line of power flow.

mined by the size of guide and the coupling. The two are connected through a hole as shown. The power flows *back* along the acute angle. The reason for this (and for directional couplers in general) is not easy to see. It depends on the fringing field at the hole and the new boundary conditions imposed by the second guide at the angle chosen.

[12] J. W. Severinghaus, Radiation Laboratory Report 55, Feb. 13, 1945.

The second is the two-hole coupler. The two holes are spaced a quarter wavelength apart, and the power flows in the original direction. This type of coupler is usually explained as due to amplitude summation and cancellation on account of the quarter wave spacing. A wave which reverses its direction introduces a change of half a wavelength between the two holes, whereas a wave which continues in the same direction keeps its phase the same. The opposition of phase in the reversed case cancels the field and removes the wave. This is an oversimplified version, but it gives the correct direction of flow.

The third is the Schwinger type, which is more involved and, although it has not been put into use very much, it is likely to be the most satisfactory.

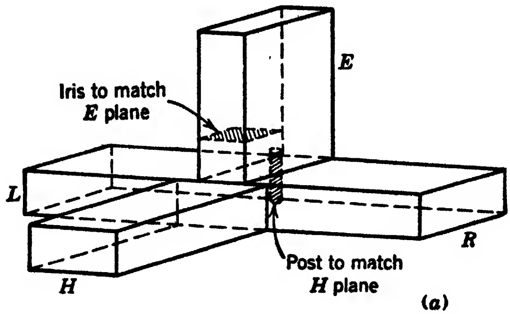
In all these couplers the power needs to be absorbed or it will return to the line and cause mismatch. This absorption is done by means of a semiconducting strip placed across the short side of the guide, so reducing it below cutoff, and cutting it like a wedge, making the transition gradual.

Magic Tees

It has been pointed out that the two kinds of tee, H plane and E plane, produce different phase relationships in the divided radiation. This fact is made use of in the "magic tee" which is a composite of the two kinds of tee, as illustrated in Fig. 4·26. The important property of the magic tee lies in the fact that it enables an asymmetric effect to be observed and measured. Thus if we consider power fed in from E in the E plane, as indicated in Fig. 4·26, it will divide equally into branches L and R, and the guide H in the H plane will have part excitation by downward field arrows as drawn and part by upward so that there will be no net excitation. Now suppose that a reflection due to some mismatch at L occurs but that none occurs at R. There is then a wave traveling all the way down the line LR with downward arrows, which is our representation of a wave from left to right. This wave gives a consistent excitation to the H plane guide H, and power flows into it because of this reflected wave only. The magic tee is therefore a directional coupler of clean and understandable design.

A simple magic tee introduces considerable reflections. These do not affect its internal operation but cause a mismatch in the

line to which the tee is attached. Thus there is, of course, some reflected power back along the E branch. In fact this corresponds to a reflection coefficient of about 40 per cent. In the H plane branch it can be as much as 60 per cent. These reflections must be



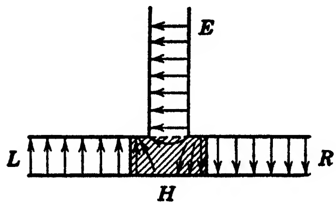


FIG. 4·26 Magic tee. A double tee is constructed as shown in (*a*). Line H is not excited by power fed in from E because the field at H is half up and half down. If an asymmetry, for example, due to a reflection at L exists, there is a net up field, and line H is excited proportionally to this asymmetry.

reduced by some means which does not destroy the symmetry. Pound [13] suggests for a 3-centimeter tee a post in the position shown to match the H plane and an iris, also indicated, to match the E plane. Magic tees have interesting applications, some of which will be considered in Chapters 8, 12, and 13. The original idea is due to Tyrrell [14] of the Bell Telephone Laboratories. The operation and possible applications of magic tees have been described by Dicke and Kyhl.[15]

[13] R. V. Pound, Radiation Laboratory Report 662, August 1945.

[14] W. A. Tyrrell, "Hybrid Circuits for Microwaves," *Proc. I.R.E.*, **35**, 1294 (1947).

[15] R. H. Dicke and R. L. Kyhl, Radiation Laboratory Report 482 (1945).

The important feature of the magic tee as just described is that, when the excitation in the arms L and R is equal in every respect, there is no excitation at all in H. For very accurate frequency discrimination this is important. However, the matching post and iris will not permit high power in the guide without break-

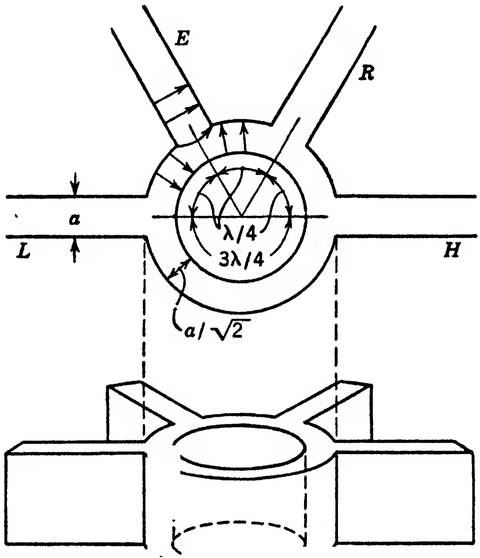


FIG. 4·27 Schematic drawing of a rat race, a device which divides power fed in at E between branches L and R, but not H. It is broadband and can support high power.

down. Therefore if such a device is to be used at high power it needs modification. This modification is shown in Fig. 4·27, and is called, reasonably enough, a "rat race."

The operation again depends on the phase division of an E plane tee. This division, which occurs at the entry E, results in two waves having the same direction across the guide at H, which is a field configuration which will not excite H. The power therefore divides equally between L and R. The same is true for power fed in at H which divides in such a way as to build up a field at E which will not excite E. The power then goes equally to L and R.

The advantage of such a rat race over a magic tee is that a match can be obtained for both E and H lines by the correct choice of the width of the circular path. The a dimension is made $1/\sqrt{2}$ times smaller than the normal guide. Since no power goes into H and it is situated an even number of half wavelengths from E, it has no effect on the impedance at E. The impedances due to L and R add at E, and so each must appear as ½ with the line E as 1. If x is the impedance of the curved guide we have, by the familiar quarter wave relation,

$$x^2 = \frac{1}{2} \times 1 \quad \text{or} \quad x = \frac{1}{\sqrt{2}}$$

This accordingly determines the width of the guide in the curved part, for the impedance of a guide is proportional to its depth.

Rat races can be made to give an attenuation of 50 decibels between power in at E and power out at H for the center of the band, and 30 decibels 10 per cent off frequency. They have no sharp edges and can support high power.

4·7 MEASUREMENT OF THE Q OF CAVITIES

Of increasing importance in microwave research is the measurement of the Q of cavities. It is used as part of the design of a microwave accelerator or in the measurement of small absorption coefficients (Chapter 13). It is necessary in the development of TR boxes.

The direct measurement of the frequency-versus-power-absorbed curve naturally enables the value of Q to be found. It is usually easy to vary frequency, and so the procedure adopted is to observe the power transmitted or absorbed at the peak (frequency f) and then the frequency at which the same quantity is reduced to half of its value. Two values, f' and f'', can be obtained on each side of maximum. Q is then by definition $f/(f'' - f')$.

Since Q is a measure of the power stored compared to the power lost, anything which increases the power lost will affect the value of Q. Thus if a load is applied to the cavity the value of Q changes. The value of Q for a matched output load is called the loaded Q. The unloaded Q is naturally greater. In order to derive the value

of the unloaded Q from the measurement of the loaded Q it is necessary to observe the power reflected by a cavity into which radiofrequency is being fed and also the position of the standing wave minimum as a function of frequency. If the cavity is highly efficient, so that the unloaded Q is much greater than the loaded Q, the measurement is simplified. The power reflected is measured as a standing wave ratio r, at resonance. Then

$$\text{Unloaded } Q = (1 + r)(\text{loaded } Q)$$ [16]

Other methods of measurement are in use. If a sensitive receiver and a pulsed r-f source are available the Q can be measured in terms of the ringing time of the cavity. Since the rate of loss of power (see page 48) is given by $e^{-\omega t/Q}$, the time to lose power after the cavity has been excited can be made a measure of Q. This is suitable in a laboratory which is equipped for radar because the pulsed r-f source and sensitive receiver are automatically present.

It is possible to measure the energy stored by means of a series of thermistors placed at various points in a large cavity and averaging the values read. This can then be compared to the power fed in (and also lost) and a direct measurement of Q made. This method enables very sensitive absorption determinations to be carried out.

Summary

This chapter can hardly be summarized. It is more to the point to consider the state of the art and to predict future progress. It is likely that the application of careful design to space charge amplifiers and oscillators will bring tubes which compete with the magnetron in output power and permit the use of amplification techniques. These will give a flexibility to the use of microwaves which is at present lacking. The present stage in design is one in which microwaves are about as well understood as circuits at lower frequencies, and the thinking is based on the knowledge gained from that field. The stage of design in which microwaves are the basis and are not treated by analogy with other circuits has just about begun. Developments like magic tees and directional couplers are products of purely microwave theory.

[16] E. Nunker, H. C. Early, G. Hok, and G. R. Bridgeford, in *Very High Frequency Techniques*, McGraw-Hill Book Co., 1947, p. 618.

The important question of limiting wavelength is hard to answer. There is no doubt that means for obtaining reasonable powers at ½ centimeter now exist. A radical change in principle may be required to enable the production of 1-millimeter waves. The technique for handling high power at such wavelengths would be quite different from anything now used unless the art of controlling the modes in a large waveguide could be brought into being.

There is a lot to do with what we now have. As a tool in physical and chemical investigation microwaves have hardly begun to be used. Some of the possibilities are being realized, and the preliminary results of such work are described in the last three chapters of this book.

PROBLEMS

4·1 Compute the radius of a cylinder which will give 100 db attenuation in 10 cm length at 10 cm wavelength in air. TE_{11} mode.

4·2 A coaxial line of outer diameter ⅝ in. with an inner conductor of diameter ⅛ in. has the inner conductor changed in diameter to ¼ in. What is the standing wave ratio in the first section of line?

4·3 Using the figures given on p. 103 construct an impedance and admittance circle chart suitable for your own use.

4·4 Show how to match the two lines above with a quarter wave sleeve at 10 cm wavelength. Calculate the standing wave ratio at 11 cm for the sleeve you propose.

4·5 Repeat the above with two adjacent sleeves a quarter wavelength long.

4·6 A coaxial line ⅝ in. outer and ½ in. inner conductor diameter is observed to break down when 500 kw are fed into it while a standing wave ratio of 3 in voltage is present. What power will it carry when properly matched?

4·7 Calculate the attenuation in 100 ft of the above cable at 10 cm.

4·8 A bend in 1½ by 3 in. waveguide is observed to give a standing wave ratio of 1.01 in voltage at 10 cm. Suggest the depth and position of a matching screw to eliminate this.

4·9 Calculate the cutoff wavelengths in a round pipe of 1 cm radius

4·10 Calculate the power-carrying capacity of ½ by 1 in. guide at 3 cm and 4 cm wavelengths.

4·11 What standing wave ratio is introduced by a capacitive symmetrical diaphragm placed in a 1½ by 3 in. waveguide with a gap of 1 in.?

4·12 Compute the gain and sidelobe values for a 10-ft paraboloid illuminated at 3 cm wavelength by uniform illumination.

4·13 Plot a graph showing how the direction of beam in a broadside array in a guide ½ in. deep changes as a is varied. Assume $\lambda = 3$ cm.

CHAPTER 5

PULSE CIRCUITS

The development of television and radar has given great impetus to the study of circuits for the generation, amplification, and detection of short voltage and current pulses of durations measured in microseconds. Such pulses are called *video pulses* since they first became important in the picture-transmission problem. Pulse techniques, particularly as developed for radar applications, have also proved to be very useful in scientific work, and it is safe to predict that their usefulness will be much extended in the near future.

5·1 SPECTRUM OF A REPETITIVE PULSE WAVEFORM

A consideration of the spectrum of frequencies contained in a video voltage pulse makes it clear why problems arise in pulse

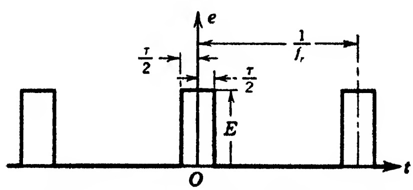


FIG. 5·1 A repetitive pulse waveform.

circuits which are of less importance in circuits designed for use at audio- and low radiofrequencies. Suppose the voltage varies with time t according to the diagram in Fig. 5·1; that is, it has the

value zero except during pulses of width τ seconds and height E volts, occurring at intervals of $1/f_r$ seconds, where f_r is the repetition frequency. Since this waveform is repetitive it can be expanded in a Fourier series and, since it is an even function of t, only cosine terms appear in the expansion:

$$e(t) = E\tau f_r \left(1 + 2\sum_{k=1}^{\infty} \frac{\sin \pi k\tau f_r}{\pi k\tau f_r} \cos 2\pi k f_r t\right)$$

$$= A_0 + \sum_{k=1}^{\infty} A_k \cos 2\pi k f_r t \qquad (5\cdot 1)$$

Thus the repetitive pulse waveform is composed of a d-c term plus all the harmonics of the repetition frequency; the harmonics

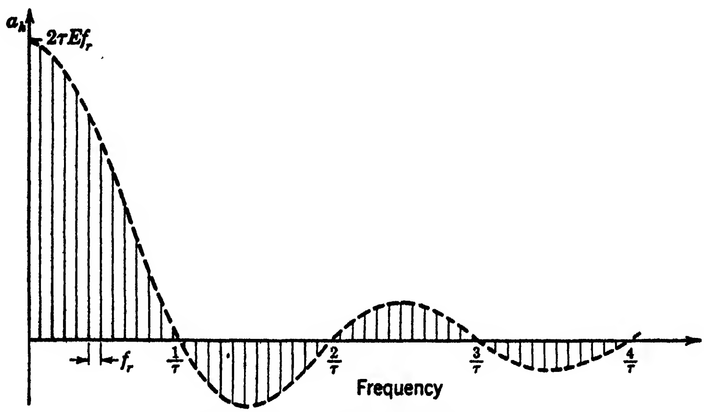


Fig. 5·2 Amplitude spectrum of pulses of width τ sec occurring at intervals of $1/f_r$ sec.

are all in phase at $t = 0$. The quantity τf_r appearing in the d-c term is called the *duty cycle*. The d-c term is thus equal to the duty cycle times the pulse amplitude. The *amplitude spectrum* of the waveform is obtained by plotting the amplitudes of the various components against their frequencies. The function $(\sin x)/x$ has zeros at $x = n\pi (n = 1, 2, 3, \cdots)$, and extrema at the roots of $\tan x = x$. Harmonics with frequencies in the neighborhood of n/τ have small amplitudes; the amplitudes have alternately positive and negative values about these zeros. These facts are illustrated in Fig. 5·2. If the pulse width is 1 microsecond, the first zero will occur at 1 megacycle.

It is evident that as the repetition frequency becomes smaller the spacing of the harmonics becomes smaller. In the limit, for $f_r = 0$, the Fourier series analysis breaks down since the waveform is no longer periodic. In this case, the amplitude and phase spectra of the waveform may be computed by the Fourier integral method (Appendix 1). It is found that, for a single rectangular pulse, the amplitude spectrum is a continuum of frequencies extending from $-\infty$ to $+\infty$, having a $(\sin x)/x$ envelope as in Fig. 5·2, with zeros occurring at $f = \pm(n/\tau)$ $(n = 1, 2, 3, \cdots)$. The maximum amplitude, at $f = 0$, is $\tau E/2\pi$.

Circuit Bandwidth

It is evident that a circuit which is required to respond to short pulses must be able to accommodate a wide range of frequencies. In particular, a video pulse amplifier must have a reasonably flat amplitude response up to high frequencies. In addition, its phase response must not deviate too greatly from the ideal relation $\phi = af$, where ϕ is the phase angle and a is a constant. However, there are very good reasons for not making the response of an amplifier any broader than necessary for the particular application for which it is designed. As will be seen in Chapter 8, unavoidable noise power in the output of an amplifier increases in proportion to the bandwidth of the amplifier. Furthermore, as shown later in this chapter, the gain which is realizable per stage in an amplifier is in general inversely related to the bandwidth, so that for a given overall amplification a broadband amplifier requires more stages than one with a narrower pass band. It thus becomes a matter of some importance to be able to judge what bandwidth is optimum for a particular application. This can only be done roughly, since different applications call for different pulse output characteristics. In radar practice, for example, it is important to balance bandwidth (with accompanying noise) against the visibility of small pulses in the presence of noise (i.e., to maximize the signal-to-noise ratio in terms of signals and noise presented on a cathode ray tube, Chapter 6), whereas in the measurement of the velocity of transmission of ultrasonic disturbances by pulse techniques the primary objective is to obtain amplified pulses with steep leading edges.

The effect of bandwidth on the ability of an amplifier to reproduce a short pulse is illustrated [1] in Fig. 5·3. Any amplifier can be looked upon as a filter, since it does not amplify signals of all frequencies. The curves in Fig. 5·3 (*a*) give the amplitude char-

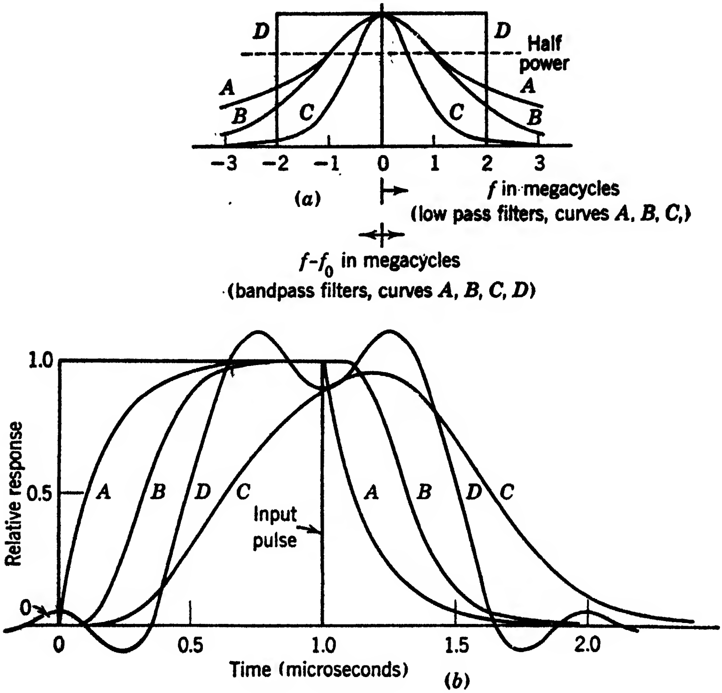


Fig. 5·3 Pulse response of various filters. (H. Wallman, Radiation Laboratory Report 285, June 1942.) (*a*) Amplitude characteristics of various low pass and bandpass filters; (*b*) response to a 1-μsec rectangular video pulse (low pass filters) or modulation pulse (bandpass filters) of filters having the characteristics shown in (*a*). *A*, single-tuned filter; *B*, *C*, 6 single-tuned cascaded filters; *D*, ideal bandpass filter (phase shift taken equal to π rad per mc).

acteristics of some important low pass and bandpass filters (curves *A*, *B*, and *C*), and the ideal bandpass filter (curve *D*). The right half of curve *A* is essentially the characteristic for a single video amplifier stage of the type shown in Fig. 5·4; it has a half-power cutoff frequency of 1 megacycle (we pay no attention here to the

[1] H. Wallman, Radiation Laboratory Report 285, June 1942.

decrease in gain at very low frequencies since it is of no importance in connection with 1-microsecond pulses). Both halves of curve A constitute the characteristic of a single-tuned amplifier using a parallel resonant coupling filter (cf. page 226), tuned to the frequency f_0 and 2 megacycles wide at the half-power frequencies. Curves B and C correspond to six video stages, or six tuned stages, in cascade, with overall bandwidths, for the video case, of 1 and ½ megacycle respectively. Figure 5·3 (*b*) shows the response of these various filters to a 1-microsecond input pulse. It will be noted that a bandwidth of even ½ megacycle (in the low pass case) is sufficient to allow the output pulse to reach nearly full amplitude, but that the pulse edges are less steep than when the bandwidth is 1 megacycle. Six cascaded stages give an output very similar to a single stage of the same bandwidth, but with a larger delay time. The square characteristic of the ideal band-pass filter gives an output which has very little advantage, regarding steepness of edges, over single-tuned stages of the same bandwidth, and has the disadvantage of causing considerable overshoot. Overshoot is invariably encountered with actual filters having sharp cutoff because of the bad phase shifts which accompany such cutoff.

A generalization of some value in the present connection is that, for several types of video amplifiers, the pulse output corresponding to a rectangular pulse input has a *rise time* τ_r related to the high frequency f'' at which the power gain is down 3 decibels by the expression

$$\tau_r \approx \frac{0.35}{f''} \tag{5·2}$$

The rise time is defined as the time in seconds required for the pulse to rise from 0.1 to 0.9 of its final amplitude.

5·2 VIDEO AMPLIFIERS

A video amplifier is a circuit which is capable of amplifying video pulses without serious distortion. The amount of distortion which can be tolerated depends entirely on the purpose for which the amplifier is to be used. This point will receive further attention below.

It is evident that video amplifiers are of central importance in television and radar. The signals obtained from a television cam-

era consist of very weak video pulses which must be amplified before they can be used to modulate the transmitter. The detected signals obtained from a radar or television receiver require amplification before they can be displayed on a cathode ray tube.

A discussion of pulse circuits might more logically start with pulse generators than with pulse amplifiers. However, it is convenient to reverse this order, since the fundamental principles of pulse circuits are more clearly brought out in connection with amplifiers, and many pulse generators contain video amplifiers.

Gain-Bandwidth Product

Most video amplifiers are of the resistance-capacitance type.[2] Consider the stages shown in Fig. 5·4. C_S is composed of the output capacitance of V_1, the input capacitance of V_2, and the stray capacitances of the wiring and sockets, all in parallel. At the high frequencies involved in the rapid rise and fall of video signals, the approximate equivalent circuit is as shown in Fig. 5·5 (*a*), since the impedance of C_C is negligible. If $R_G \gg R_L$, the ratio of the voltage gain A_H at high frequencies to the gain A at moderate frequencies (where the impedance of C_S is large compared to R_L) is

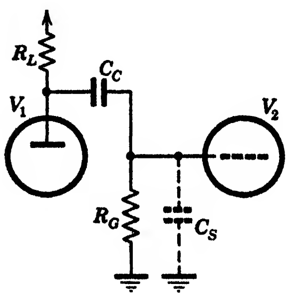


FIG. 5·4 Schematic diagram illustrating resistance-capacitance coupling of two amplifier stages.

$$\frac{A_H}{A} = \frac{1}{1 + 2\pi jfR_LC_S} \tag{5·3}$$

where f is the frequency in cycles per second. Therefore the frequency f'' at which the output power is down 3 decibels from the midband power, with a given input amplitude, is given by

$$f'' = \frac{1}{2\pi R_LC_S} \tag{5·4}$$

For a pentode, since $r_p \gg R_L$ in any practical video amplifier,

$$A = \frac{\mu R_L}{r_p + R_L} \approx g_mR_L \tag{5·5}$$

[2] F. E. Terman, *Radio Engineers' Handbook*, McGraw-Hill Book Co., 1943, pp. 413 ff.

where r_p, μ, and g_m are respectively the plate resistance, amplification factor, and grid plate transconductance. The product $f''A$, called the *gain-bandwidth*[3] *product,* is seen to be independent of R_L, and thus serves as a convenient figure of merit for a tube serving as a video amplifier since the minimum value of C_S is largely determined by the tube type. Values of the gain-bandwidth product for two important tube types are given in Table 5·1. Reference to this table shows, for example, that a 6AC7 stage with a voltage gain of 10 (R_L = 1100 ohms) cannot have a

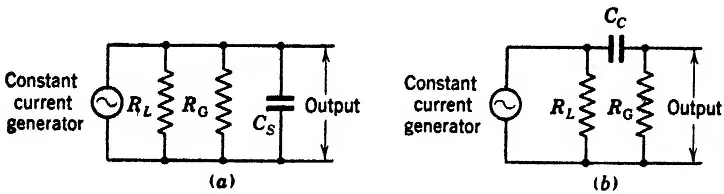


FIG. 5·5 Equivalent circuits for resistance-capacitance coupling. (*a*) High frequency equivalent circuit; (*b*) low frequency equivalent circuit.

bandwidth exceeding 5.7 megacycles, unless compensating networks are employed.

TABLE 5·1 CHARACTERISTICS OF TUBES SUITABLE FOR USE IN VIDEO AMPLIFIERS

Tube Type	Transconductance, μmhos	C_S, μμf	Gain-Bandwidth Product
6AC7	9000	25	57
6AK5	5000	12	66

High Frequency Compensation

The value of f'' can be somewhat increased without decreasing the midfrequency gain by various schemes which compensate for the loss in high frequency gain due to the capacitance C_S. As a simple example, consider the case in which it is required to supply self-bias to V_1 (Fig. 5·4) by means of a cathode resistor R_K. If this resistor is not bypassed, the cathode degeneration will reduce the gain at all frequencies by the factor $1/(1 + g_mR_K)$. If R_K is

[3] The high frequency 3-db point is practically equal to the 3-db bandwidth in a video amplifier. A further discussion of the gain-bandwidth product will be found in Chapter 7.

small it is difficult to use a large enough bypass condenser to remove the degeneration at very low frequencies. It can be shown (Problem 5·3) that A and A_H/A retain the values given by equations 5·5 and 5·3 if the load resistor is increased to $R_L(1 + g_mR_K)$ and R_K is bypassed by a condenser $C_K = (R_L/R_K)(1 + g_mR_K)C_S$. The incomplete removal of the degeneration obtained in this way thus allows a larger load resistor to be used without decrease in bandwidth, though of course with no increase in midfrequency gain.

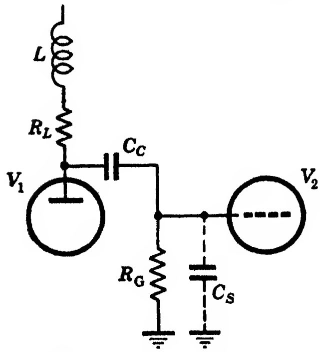


FIG. 5·6 Schematic diagram of the shunt-peaking method of high frequency compensation.

Shunt Peaking

If a small inductance is placed in series with the load resistor of V_1, as in Fig. 5·6, the high frequency response can be considerably improved. Since the inductance is essentially in parallel with C_S it is spoken of as a *shunt-peaking* coil. The impedance of the coil increases at high frequencies, thereby tending to compensate for the loss of gain due to the capacity C_S. If we again assume that $r_p \gg R_L$ and $R_G \gg R_L$, we find that the ratio A_H/A has the value

$$\frac{A_H}{A} = \frac{1}{R_L} \frac{R_L + j\omega L}{1 - \omega^2 LC_S + j\omega R_L C_S} \tag{5·6}$$

Rationalization of this expression, and substitution of the quantities

$$m = \frac{L}{C_S R_L{}^2} \tag{5·7}$$

and

$$\omega_0 = \frac{1}{C_S R_L} \tag{5·8}$$

gives the results

$$\left|\frac{A_H}{A}\right|^2 = \frac{1 + m^2 \dfrac{\omega^2}{\omega_0{}^2}}{1 - (2m - 1)\dfrac{\omega^2}{\omega_0{}^2} + m^2 \dfrac{\omega^4}{\omega_0{}^4}} \tag{5·9}$$

$$\tan \phi = (m - 1)\frac{\omega}{\omega_0} - m^2 \frac{\omega^3}{\omega_0^{\,3}} \tag{5·10}$$

where ϕ is the phase angle. Values of the relative response and phase angle for various values of the parameter m are plotted in Figs. 5·7 and 5·8.

For proper amplification of a pulse it is important that the amplitude response of the amplifier be constant to as high a fre-

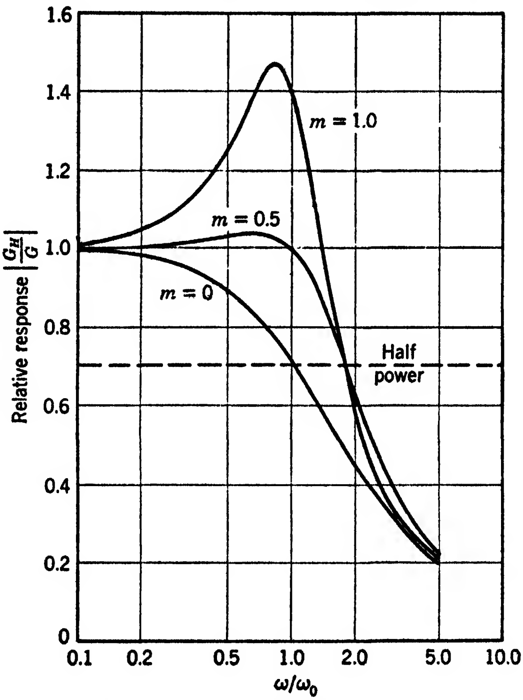


FIG. 5·7 Relative response of a shunt-peaked coupling network. The 3-db bandwidth is increased approximately 80 per cent with $m = 0.5$.

quency as possible. It is also important that the *delay time* of the amplifier remain reasonably constant to as high a frequency as possible, since distortion will result if the various components of the pulse spectrum arrive at the output terminals at different times. The delay time is given by $\tau = \phi/\omega$, so that for τ to be constant it is necessary that ϕ be proportional to ω. An amplifier with $m = 1$ would not be satisfactory because of the high frequency

peak in gain and the deviation of the phase angle from linearity. The response of such an amplifier to a rectangular pulse input would be a pulse with considerable overshoot at its leading edge and undershoot at its trailing edge. A value of m of about 0.5 gives satisfactory results.

Because of the importance of the phase response of an amplifier it is unsafe to judge a broadband amplifier by the flatness of its amplitude response, as is usually done with audio amplifiers.

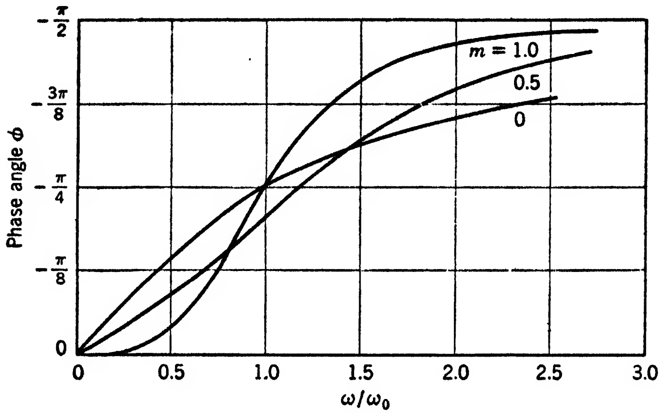


FIG. 5·8 Phase angle of a shunt-peaked coupling network. With $m = 0.5$ the phase angle is nearly linear with frequency over the pass band of the network.

Elaborate coupling networks can be devised which give a flat amplitude response over a very broad band, but accomplish this at the expense of very rapid high frequency cutoff accompanied by non-linear phase shift. Such an amplifier would be less satisfactory than one with a narrower amplitude response but better phase characteristics. In actual practice the best check on the high frequency response of a video amplifier is observation of the output produced by an input rectangular pulse. The input pulse should have rise and fall times short compared to the expected output. For simple coupling networks such as that shown in Fig. 5·6 a rough correlation between the output rise time and the 3-decibel cutoff frequency, f'', is given by equation 5·2.

With $m = 0.5$ the 3-decibel frequency f'' is about 1.8 times the value with no compensation. Thus a shunt peaked ($m = 0.5$) 6AC7 stage having a gain of 10 should have $f'' \approx 10$ megacycles.

In some cases it may be desired, for pulse-shaping or other purposes, to operate a video amplifier tube outside the region of signal amplitudes where its characteristics are approximately constant. When this is done the above treatment is not applicable. Thus, if a positive pulse is applied to the grid of a stage biased beyond cutoff, the leading edge of the output pulse will be steeper than the trailing edge if the load resistor exceeds a few hundred ohms. This is because the capacity C_S can be discharged through the low impedance of the tube during conduction more rapidly than it can be recharged through the load resistor.

Further discussion of coupling networks which may be employed in video amplifiers to secure high frequency compensation can be found in several places.[4] The simple shunt peaking scheme is sufficient for many applications.

Other Methods of High Frequency Compensation

Numerous methods of high frequency compensation have been employed. For example, the effect of the distributed capacity can be overcome to some extent by adding to it an equal negative

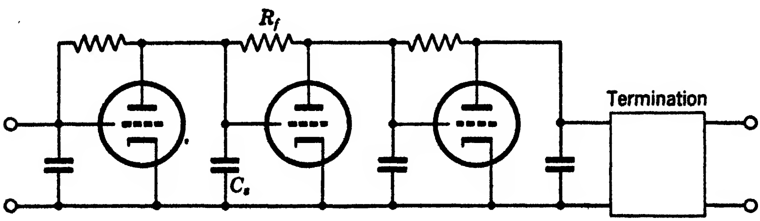


FIG. 5·9 Schematic diagram of a feedback chain, omitting blocking condensers and d-c feeds.

capacitance. This *neutralization* procedure has been much employed in the past in reducing the large input capacitance of triodes resulting from their grid plate capacity. H. L. Schultz of Yale University has shown that this type of compensation is useful with pulses supplied from a high impedance source such as an

[4] F. E. Terman, *Radio Engineers' Handbook*, McGraw-Hill Book Co., 1943, pp. 418 ff.; V. K. Zworykin and G. A. Morton, *Television*, John Wiley and Sons, Inc., 1940, pp. 405 ff.; H. E. Kallman, R. E. Spencer, and C. P. Singer, *Proc. I.R.E.*, 33, 169 (1945).

ionization chamber. The negative capacitance is furnished, in effect, by *regenerative* feedback; it is essential that the regenerative signal be supplied by a highly stable *degenerative* amplifier stage.

Ferguson [5] has discussed in some detail the use of *feedback chains* in broadband amplifiers. Figure 5·9 is a schematic diagram (omitting blocking condensers, and so forth) of such a circuit. Ferguson has investigated the response obtained with various types of terminations for the chain.

Low Frequency Response

The low frequency equivalent circuit of a simple video amplifier is given in Fig. 5·5 (*b*). Loss in gain occurs chiefly as a result of increase in the impedance of the coupling condenser C_C. It can be shown that

$$\frac{A_L}{A} = \frac{2\pi j f C_C R_G}{1 + 2\pi j f C_C R_G} \qquad (5\cdot11)$$

where A_L is the voltage gain. The frequency at which the power is down 3 decibels is

$$f' = \frac{1}{2\pi C_C R_G} \qquad (5\cdot12)$$

The method usually employed to secure adequate low frequency response is to use a large coupling condenser and a large grid resistor. Thus if $C_C = 0.1$ microfarad and $R_G = 500K$, $f' = 3$ cycles.

If necessary, low frequency compensation can be obtained by means of the resistor-condenser combination shown in Fig. 5·10. At low frequencies the impedance of C_1 increases, so that the effective load resistor also increases. Good compensation is secured when $C_1 R_L = C_C R_G$ and $R_1 \geqq 10(C_C R_G / C_1)$, provided the impedance of the power supply is low.

We saw above that the response of an amplifier to the high frequencies contained in the leading and trailing edges of a pulse gives the most satisfactory indication of the merit of a video amplifier at high frequencies. Correspondingly, the low frequencies

[5] A. J. Ferguson, *Can. J. Research*, **A24**, 56 (1946).

in the flat top of a square wave are useful in checking the low frequency response. The frequency of the square wave should be of the order of the cutoff frequency of the amplifier; 60 cycles is a convenient frequency for many cases. The amount of drop in the top of the output square wave which can be tolerated depends on the use to be made of the amplifier. For a single stage, in the absence of compensation, the amplitude will drop to $1/e$ of its original value in $C_C R_G$ seconds, provided the power supply has adequate regulation and there is no time constant of this order of magnitude in the screen and cathode circuits. Overcompensation at low frequencies, or coupling between stages through the power supply impedance, or insufficient power supply regulation may cause the overall gain to rise more or less sharply at low frequencies; this is indicated by a rise in the top of the output square wave, at least during the first part of the cycle. Such a situation may lead to a very low frequency oscillation called "motorboating." In order to reduce the likelihood of motorboating, which is particularly apt to occur with video amplifiers, the response should not extend to lower frequencies than necessary.

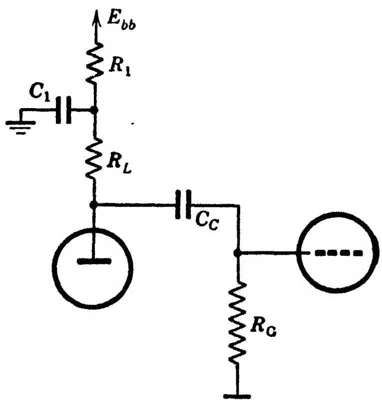


FIG. 5·10 Illustrating a method for low frequency compensation in an amplifier. Because of the increase of the impedance of C_1 at low frequencies, the effective load impedance increases to counterbalance increase in the impedance of C_C.

If a screen-dropping resistor and bypass condenser are used, there will be a decrease in gain at low frequencies if the condenser is not large enough, because of the introduction of screen degeneration. Another source of decrease in gain at low frequencies is the bypassing of the cathode resistor by a condenser if self-bias is used. Terman [6] discusses means for compensating for these degenerative effects.

[6] F. E. Terman, *Radio Engineers' Handbook*, McGraw-Hill Book Co., 1943, pp. 413 ff.

Cascaded Stages

If two stages having 3-decibel frequencies f' and f'' are cascaded, the overall gain will be down 6 decibels at f' and f''. Thus the bandwidth of the cascaded pair will be considerably less than that of the individual stages. The exact relation between the overall bandwidth and the number of cascaded stages depends on the type of amplitude response curve for the individual stages. For an amplifier made up of n stages, each one of which may be described by the equivalent circuits of Fig. 5·5, the low and high 3-decibel frequencies are given by

$$f'_n = \frac{f'}{\sqrt{2^{1/n} - 1}} \tag{5·13}$$

$$f''_n = f''\sqrt{2^{1/n} - 1} \tag{5·14}$$

5·3 VIDEO PULSE GENERATORS

It will be convenient to divide the discussion of video pulse generators into two parts, the first concerned with the formation of low level pulses and the second, with that of high level pulses. Pulses with amplitudes in excess of a few hundred volts will arbitrarily be classed as high level pulses.

The quality of the output from a pulse generator is of considerable importance. For some purposes, such as the modulation of a magnetron (Chapter 3), rather rigid pulse specifications must be laid down. In this particular application, since the magnetron can oscillate in unwanted modes at the wrong voltage, both the rise time [7] and the fall time need to be short, and the pulse top needs to be fairly flat; in addition no large oscillations should follow the pulse. In other applications the pulse specifications can be considerably less rigid. For example, a trigger pulse (that is, a pulse used for synchronizing the operation of various circuits) is usually specified only in regard to the steepness of its leading edge and its amplitude, though in some cases its width, measured at,

[7] The rise time customarily refers to the leading edge of a pulse regardless of the sign of the pulse. The modulation pulse applied to a magnetron is usually negative, since it is convenient to have the magnetron anode at ground potential.

say, half amplitude, is significant. In other words, a well-designed triggered circuit should not be sensitive to the shape of the triggering pulse.

It should be noted that in general the output developed by a pulse generator is dependent on the characteristics of the load into which it works. Some loads, such as a magnetron, present an impedance which changes during the pulse interval. It is thus evident that a fully significant statement of the characteristics of the pulse output can be given only if the load is carefully defined.

Low Level Pulse Generators

A great variety of circuits have been developed for producing short pulses with amplitudes up to a few hundred volts. We will limit our discussion to four important types, namely blocking oscillators, multivibrators, gas tube circuits, and shaping circuits.

(*a*) *Blocking Oscillators.* The circuit of a typical *free-running* blocking oscillator is shown in Fig. 5·11. There is very close re-

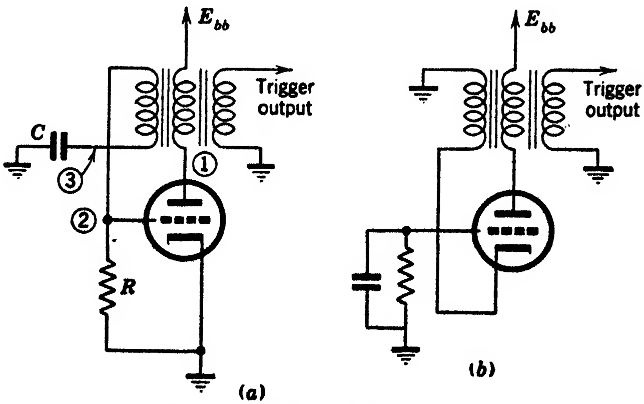


FIG. 5·11 Free-running blocking oscillators with (*a*) grid coupling, (*b*) cathode coupling. The numbers in (*a*) refer to the waveforms of Fig. 5·12.

generative coupling between the plate and grid by means of a pulse transformer (page 167); when oscillation starts the plate goes negative and the grid is driven positive, so that a considerable negative charge develops on the condenser C as a result of grid current. Thus, when the grid swings negative on the second half

of the cycle, it goes far beyond cutoff and remains there until the charge on C leaks off through the resistor R. This cycle of operation is repeated when the grid finally starts to go above cutoff. The voltage waveforms characteristic of a blocking oscillator are shown in Fig. 5·12, the waveforms being numbered to correspond

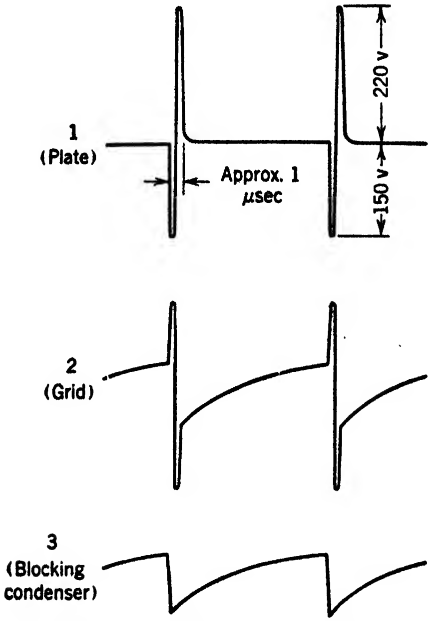


FIG. 5·12 Typical waveforms obtained with a grid-coupled blocking oscillator. The waveforms were observed at the indicated points of Fig. 5·11 (*a*), with $R = 100K$, $C = 0.01$ μf, $E_{bb} = 300$ v, a Utah X-124T-2 transformer, and one-half of a 6SN7 tube. The pulse widths are much exaggerated relative to the interpulse periods.

to the indicated positions in the circuit of Fig. 5·11. (The voltages in the second and third curves are not drawn to the same scale as in the first. The circuit giving these voltages had $R = 100K$, $C = 0.01$ microfarad, $E_{bb} = +300$ volts, and a Utah X-124T-2 pulse transformer was used, with one half of a 6SN7 dual triode.) The output from the pulse transformer is similar to the plate waveform, but reversed in phase if connected as shown.

The repetition rate of the output pulses is determined by the value of the time constant RC and the voltage to which the re-

sistor R is returned. It is evident that there will be less "jitter" in the period of the operation if this resistor is returned to E_{bb} or some other positive voltage. The circuit can be easily synchronized with an external synchronizing voltage of slightly higher frequency, applied to the grid, or the plate, or by means of another winding on the pulse transformer.

The stability of the frequency of a free-running blocking oscillator is improved by using cathode coupling instead of grid cou-

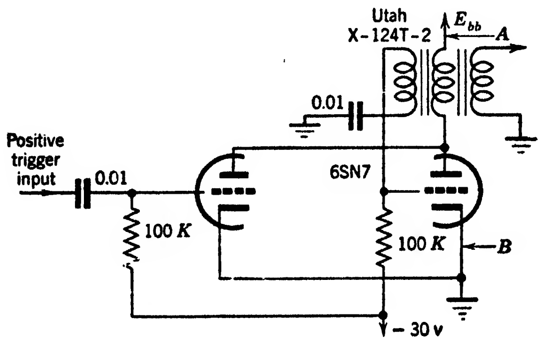


FIG. 5·13 A typical biased blocking oscillator. An output pulse is obtained only when an input trigger is supplied. Pulses free from overshoot can be obtained by inserting a small resistor ($\approx 100\Omega$) at A or B and using the voltage developed across the resistor as the output.

pling, as illustrated in Fig. 5·11 (*b*). The increase in stability results from the fact that the minimum plate-cathode potential is larger than in the grid-coupled case, so that fluctuations in the heater voltage have less effect. The plate winding should have about twice as many turns as the cathode winding to obtain proper impedance matching.

For many purposes it is desired to have the blocking oscillator quiescent except when "fired" by an external trigger. This is accomplished by biasing the grid below cutoff. A typical *biased* blocking oscillator is shown in Fig. 5·13, together with a convenient method of applying the trigger through a buffer amplifier. The output from the pulse transformer is similar to that obtained with the circuit of Fig. 5·11.

If a pulse without overshoot is desired, the blocking oscillator output can be passed through a diode or applied to the grid of an appropriately biased tube. A very useful method of obtaining

a clean pulse from a very low impedance source is to use a current rather than a voltage pulse. If a small resistor (about 100 ohms) is inserted at position *A* in Fig. 5·13, a negative pulse without overshoot will be obtained at the connection between the resistor and the transformer; if a similar resistor is inserted at position *B* a clean positive pulse will be obtained. These pulses have very steep leading edges, with somewhat sloping trailing edges. With the circuit of Fig. 5·13 they are about 1 microsecond wide and

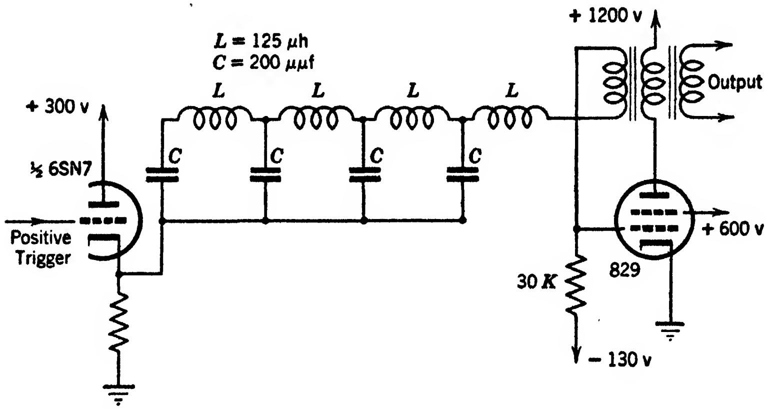


FIG. 5·14 A medium-power blocking oscillator employing an open-ended artificial transmission line for pulse shaping.

about 50 volts high, the pulse at the cathode being somewhat the larger because of the contribution of grid current. It is interesting to observe that the magnitude of these pulses shows that the tube passes about half an ampere during the pulse. Because of this fact it is frequently advisable to insert a decoupling network, consisting of a resistor with a condenser to ground, between the pulse transformer and the plate supply.

The width of the pulse obtained from a blocking oscillator is determined mainly by the size of the blocking condenser, the pulse transformer, and the tube type. A 0.1-microsecond pulse can be obtained with C = 100 micromicrofarads, a Utah X-124T-2 transformer, and a 6AG7 video power pentode. The pulse width can be closely controlled if the blocking condenser is replaced by an artificial transmission line used as a pulse-forming line (page 175). A circuit of this type is shown in Fig. 5·14. This particular circuit develops a positive pulse of about 1-micro-

second width and 1-kilovolt amplitude across a load of $2K$ impedance. A positive trigger, which must come from a low impedance source such as a cathode follower (page 298; the 6SN7 is here connected as a cathode follower), is fed through the capacity of the pulse line to one of the secondaries of the pulse transformer. This raises the grid of the 829 above cutoff and starts a regenerative process which rapidly drives the grid of the 829 positive. The flow of grid current charges the pulse line; after an interval equal to twice the delay time of the line the charging process is completed, the flow of grid current stops, and the grid potential drops. This process initiates a rapid regenerative cutting off of the 829 which is aided by the negative charge stored in the pulse line. This negative charge leaks off through the $30K$ grid resistor between pulses.

(*b*) *Multivibrators.* A triggered multivibrator is a convenient source of low level pulses of variable width. A discussion of this type of circuit is given in Section 10·3. The circuit illustrated in Fig. 10·10 can be employed for this purpose, the output being taken from either plate or either grid of V_2, usually through a buffer amplifier. If pulses with very rapid rise and fall times are required, the circuit must be modified in accordance with the principles outlined in the preceding section.

(*c*) *Gas Tube Pulse Generators.* The circuit of Fig. 5·15 uses a pulse-forming line to control the width of the pulse produced by a thyratron. A positive trigger from a low impedance source initiates the discharge of the pulse line through the low impedance of the thyratron into a resistor equal to the characteristic impedance of the line. When the pulse line is completely discharged the voltage across the thyratron falls below the extinction value. The amplitude of the pulse is approximately half the voltage to which the line is initially charged. The resistor R must be small enough so that the line is practically fully charged between pulses, but large enough so that the current flowing through it during the pulse is small compared to the pulse current.

A very useful pulse generator giving pulses of variable width and steep leading and trailing edges is shown in Fig. 5·16. This circuit was developed by G. D. Forbes of the Radiation Laboratory. Both thyratrons are normally non-conducting. When a positive trigger is applied to the grid of V_1 the potential of the

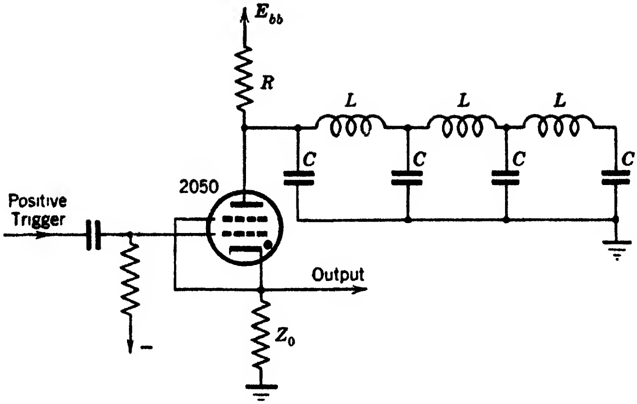


FIG. 5·15 A thyratron pulse generator using a pulse-forming line.

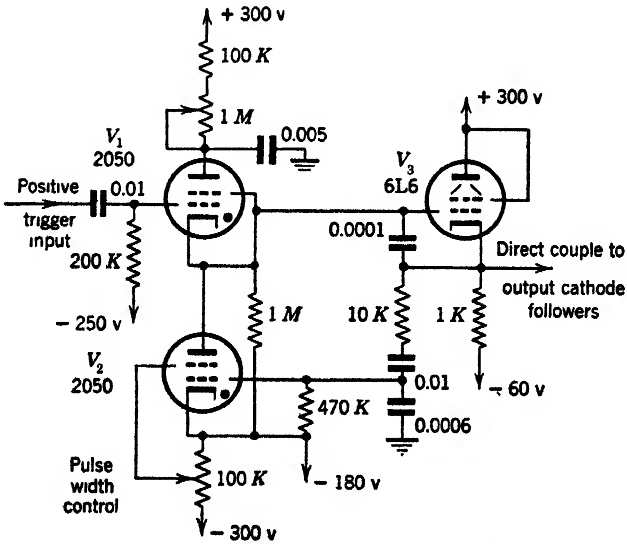


FIG. 5·16 Gas tube circuit for generating video pulses with steep edges and variable width (Forbes).

grid of the cathode follower V_3 is rapidly raised since V_1 has a very low impedance once the gas discharge has started. Within a short time the plate of V_1 becomes so close in potential to the cathode that the discharge is extinguished. The charge on the 0.0001-microfarad condenser leaks off slowly through the 1-megohm resistor, so that the potential of the grid of V_3 remains almost constant during the pulse. The feedback from the output through the $10K$ resistor gradually raises the potential of the grid of V_2 until, at a time determined by the bias of the shield grid, this tube fires and very rapidly discharges the 0.0001-microfarad condenser. V_2 is self-extinguishing just as is V_1. The 1-megohm variable resistor in the plate circuit of V_1 must be adjusted when the repetition frequency is changed. As indicated, the cathode of V_3 can be coupled directly to the grids of one or more 6L6's in parallel serving as output cathode followers. By using several tubes in parallel one can obtain pulses of considerable amplitude across a properly terminated 75-ohm video cable (see page 300).

(*d*) *Shaping Circuits.* A pulse waveform can be developed from a sine wave by the circuits of Fig. 5·17 and 5·18. The sine wave

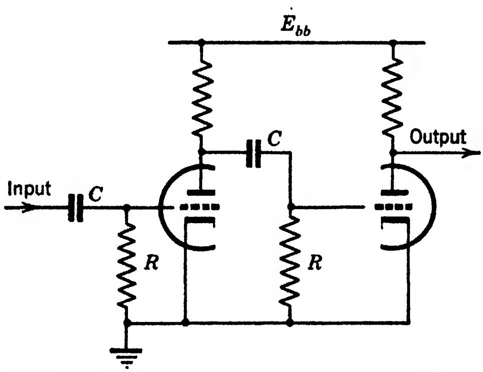


FIG. 5·17 Two-stage "squaring" amplifier for converting a sine wave into a square wave. The sharp cutoff triodes are driven to grid current on the positive half-cycles of their respective signals, and to cutoff on the negative half-cycles.

is first squared by a squaring amplifier (Fig. 5·17), which consists of two or more sharp cutoff triode stages which are driven to grid current on the positive half-cycles of their respective grid signals

and to cutoff on the negative half-cycles. The amplitude of the sine wave should be several times the cutoff voltage, and the time constants RC several times the period of the sine wave. The square wave thus produced will be more or less asymmetrical, chiefly because of the bias developed by the grid current drawn by the first grid; the asymmetry can be somewhat controlled by placing resistors in series with the grids to limit grid current.

The square wave is converted to alternate positive and negative pulses by "differentiation." The differentiating circuit is a high

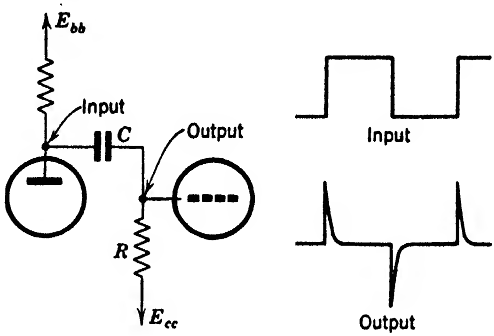


FIG. 5·18 Short time constant coupling between two amplifier stages for differentiating a square wave.

pass filter composed of a C-R combination with a short time constant, and is usually employed as the coupling between two amplifier stages (Fig. 5·18). If the square wave has very steep leading and trailing edges, and if R is large compared to the load resistor of the preceding tube, the voltage developed across R at the rise or fall of the square wave will be practically equal to the amplitude of the square wave, since the condenser has a very low impedance for the high frequencies contained in the steep edge. The output will drop to $1/e$ of its initial value in CR seconds; thus with $C = 50$ microfarads and $R = 100K$, the output pulses will have a width of the order of 5 microseconds. It should be observed that such a short time constant as this can be profitably employed only if the square wave is supplied by an amplifier stage having good high frequency response. The second tube in Fig. 5·18 may be a triode or pentode biased beyond cutoff, so that only the positive pulses will be amplified. In place of this arrangement, the

second tube can be a gas tube such as an 884 or a 2050; a large plate resistor shunted by a small capacitor to ground and a small cathode resistor across which the output is developed complete the circuit.

Differentiation may also be accomplished by means of a pulse transformer, since such transformers have good high frequency but poor low frequency response. The circuit of Fig. 5·19 illustrates a pulse transformer serving as both differentiator and blocking oscillator transformer.

It is obvious that video amplifier stages can be employed for further shaping of pulses. One of the most satisfactory ways for

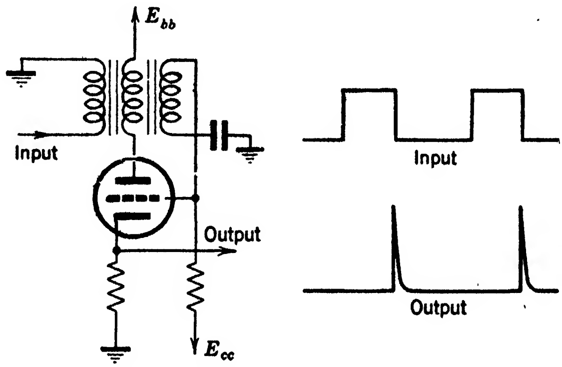


FIG. 5·19 Illustrating a pulse transformer serving as both a high pass coupling for differentiation and a blocking oscillator transformer.

producing a flat-topped positive pulse with steep edges is to drive the grid of a well-designed normally "on" video stage to cutoff by means of a relatively large negative pulse. This method involves a large current drain on the power supply if a sizable output pulse is desired; this difficulty can in some cases be avoided at the expense of additional circuit complication by appropriate pulsing of the power supply.

High Level Pulse Generators

Our discussion of pulse generators of types which can produce pulses with amplitudes up to 50 kilovolts, and peak pulse powers up to 10 megawatts, will consist essentially of a discussion of some

of the important types of circuits [8] developed for producing the modulation pulses applied to microwave radar transmitters. Since such circuits cannot in general be described without reference to the load into which they work, in most of our discussion we will

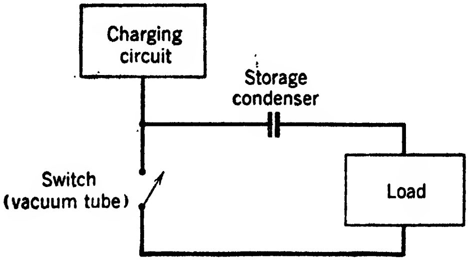


FIG. 5·20 Basic circuit of the hard tube type of high level pulse generator.

consider the load to be an appropriate microwave magnetron (Chapter 3). Such oscillators, together with their filament transformers, constitute a load having roughly the characteristics of a capacitance to ground of the order of 100 micromicrofarads,

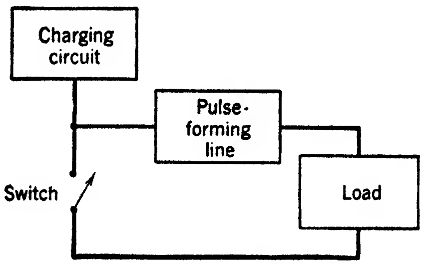


FIG. 5·21 Basic circuit of the line type of high level pulse generator.

shunted by a very high resistance except when the pulse amplitude is within about 10 per cent of operating voltage, in which case the shunt resistance drops rapidly with increasing pulse amplitude to a value of 500 to 1250 ohms, depending on the magnetron.

[8] Discussions of these circuits have been given by A. A. Jerrems and E. R. Kravitz, "Modulator Text," Radiation Laboratory Report T-15, Dec. 1943; and G. N. Glasoe and H. J. White in "General Lecture Series on Radar Components," Radiation Laboratory Report T-18, Dec. 1944. A detailed treatment will be found in the *Radiation Laboratory Technical Series*, published by McGraw-Hill Book Co.

High level pulse generators can be divided into two basic types, the *hard tube* and the *line* types. These names are derived from the fact that the former type employs high vacuum tubes as switches, whereas the latter is characterized by the use of a pulse-forming line for storing the pulse energy between pulses. Figures 5·20 and 5·21 show the fundamental circuits of these two types. In the hard tube generator energy from the charging circuit is stored in the condenser. When the switch is closed energy flows from the condenser into the load until the switch is opened again. To produce across the load a flat-topped voltage pulse with fast rise and fall times it is necessary that (*a*) the switch close *and open* rapidly and (*b*) the condenser be large enough so that the voltage across it does not decrease appreciably during the pulse. In the line-type generator, energy is stored in the capacity of the open-ended artificial transmission line. It is a property of such a line that when the switch is closed the line develops across a matched load a voltage pulse equal in amplitude to one-half the original voltage impressed on the line and of duration determined by the effective electrical length of the line. At the end of the pulse the voltage on the line, and therefore on the load, goes rapidly to zero. There is thus no need for the switch to reopen rapidly in this type of generator.

Hard Tube Modulators

It is to be noted that the hard tube type of generator is essentially a rather highly glorified video amplifier; therefore the principles discussed earlier in this chapter apply. However, in most cases, the output tube, or switch tube, is driven from below cutoff into the region where heavy grid current is drawn. Thus the switch tube behaves as an essentially infinite impedance between pulses and as a resistance of a few hundred ohms or less during each pulse.

Figure 5·22 gives the circuit of a typical medium power hard tube modulator. It is permissible here to use a large resistor in the plate circuit because during the pulse this resistor is shunted by the low resistance of the switch tube. In order for the top of the pulse applied to the magnetron to be reasonably flat the condenser C must be large enough so that the voltage across it does not decrease appreciably during the flow of the pulse current I_p.

If R_M is the magnetron impedance during the pulse and R_i is the internal resistance of the switch tube, the pulse current is given by

$$I_p = \frac{V_C}{R_i + R_M} = C\frac{dV_C}{dt} \tag{5·15}$$

where V_C is the initial voltage across the condenser. Since the pulse is very short we can replace dV_C by ΔV_C, the change in V_C

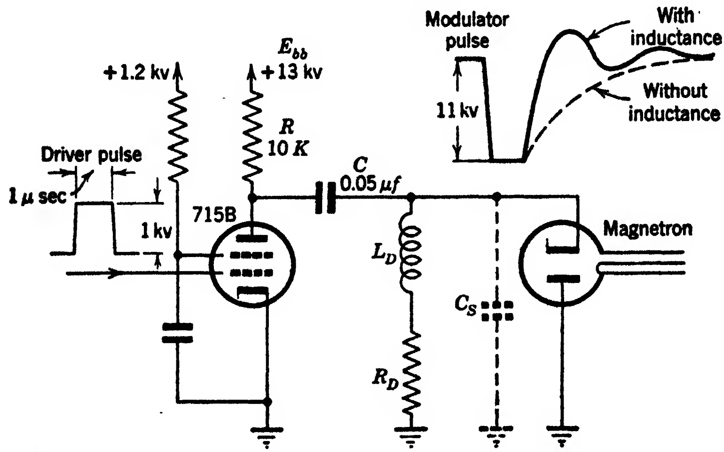


FIG. 5·22 Circuit diagram of a typical medium power hard tube modulator.

during the pulse, and dt by τ, the pulse width. Equation 5·15 then becomes

$$\frac{\Delta V_C}{V_C} = \frac{\tau}{C(R_i + R_M)} \tag{5·16}$$

We may take as typical values $\tau = 10^{-6}$ second, $R_i = R_M = 1K$. In order for $\Delta V_C/V_C$ not to exceed 0.02, C must be at least 0.05 microfarad. If C is to serve also as a filter condenser, as it sometimes does, it will have to be considerably larger.

If there is no inductance in the circuit (L_D shorted out) the voltage on the magnetron will decay with a time constant determined by the value of R and R_D in parallel and C_S, where C_S is the stray capacity. The decay of the pulse can be considerably speeded up by adding the inductance L_D so that the combination L_D, R_D, and C_S forms an approximately critically damped resonant circuit with a resonant period of the order of the pulse width. In

some modulators, $R_D = 0$, and the oscillations of the L_D–C_S circuit are damped by means of a diode with grounded cathode in parallel with the magnetron.

The switch tube of a hard tube modulator requires a large positive pulse on its grid. In low power circuits a positive pulse may be obtained from an "on" tube, but obviously in the present application the power loss in such a tube would be prohibitive. In order to avoid this difficulty, the grid signal may be supplied from an "off" driver stage through a pulse transformer connected to

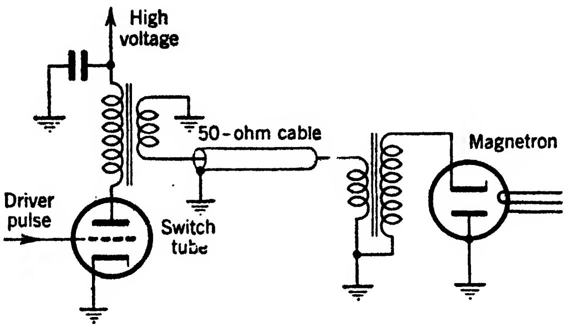


FIG. 5·23 Illustrating the use of a pulse transformer for matching a hard tube switch to a low impedance cable.

reverse the sign of the driver output pulse. For example, the switch tube in Fig. 5·22 can be driven by the blocking oscillator of Fig. 5·14.

In some applications it is important to have the magnetron located at some distance from the modulator. This can be accomplished by the use of a coaxial cable to transmit the pulse from the modulator to the oscillator, together with suitable pulse transformers to match the switch tube to the cable and the cable to the magnetron, as illustrated in Fig. 5·23.

Line Modulators

Three important types of switches [9] are used in line modulators, namely rotary spark gaps, enclosed stationary spark gaps (sometimes called trigatrons), and gas tubes such as hydrogen thyra-

[9] F. S. Goucher, J. R. Haynes, W. A. Depp and E. J. Ryder, *Bell System Tech. J.*, **25**, 563 (1946).

trons. The function of these switches is to connect one side of a charged pulse-forming line to ground through a low impedance path at specified intervals.

(*a*) *Rotary Spark Gap.* One type of rotary spark gap consists of a set of tungsten electrodes mounted on a rotating insulating disk in such a way that they pass close to a high potential

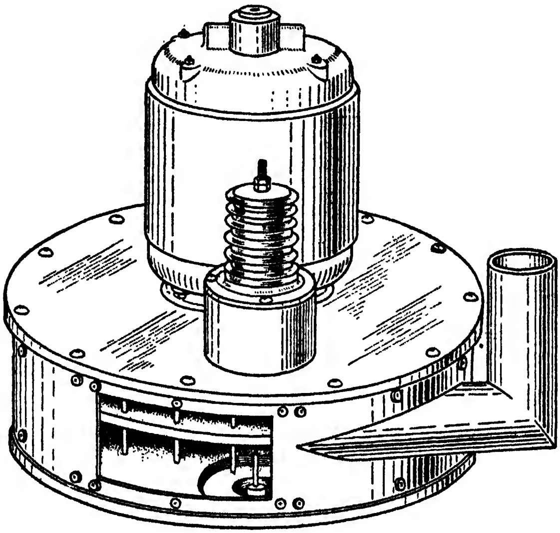


FIG. 5·24 A rotary spark gap with drive motor.

tungsten electrode connected to one side of the pulse line and a grounded tungsten electrode (see Fig. 5·24). A spark jumps from the high potential electrode to one of the electrodes on the disk, and from this disk electrode to the grounded electrode, thereby providing for a brief period a low impedance path from the line to ground. This type of switch can carry high currents and is therefore particularly useful with high power ground and ship radars. The rotary gap obviously cannot be triggered externally, so that a trigger must be taken from the modulator itself, usually from a winding on the pulse transformer coupling the pulse cable to the magnetron, for synchronizing indicators and other equipment with the modulator; that is, self-synchronous operation must be employed. For some applications this limitation, and the fact

that there is considerable jitter in the interpulse spacing, amounting to something like ±30 microseconds, preclude the use of rotary gap modulators. The pulse repetition frequency is the product of the speed of the motor which rotates the insulating disk and the number of rotating electrodes, and is therefore limited to values below about 800 pulses per second.

(*b*) *Stationary Spark Gap.* The most important form of trigatron is the so-called series gap, the name arising from the fact that two or more gaps in series must be used. Each gap consists of an anode rod surrounded by a cylindrical cathode, enclosed in a gaseous mixture such as hydrogen and argon at a pressure in the neighborhood of 1 atmosphere. Firing of the gaps is achieved by overvolting one or more of the gaps by a large external trigger; breakdown of the triggered gap or gaps applies the full pulse line voltage to the remainder of the gaps and causes them to break down also. Series gaps can be triggered with considerable accuracy, the lag of the modulator pulse after the trigger being constant within a very small fraction of a microsecond.

(*c*) *Hydrogen Thyratron.* The hydrogen thyratron finds important application in medium power modulators. It is a triode containing hydrogen gas, and it is characterized by a high peak current and a low tube drop. It is fired by a positive trigger of the order of 100 volts, and is well suited to applications requiring great accuracy of triggering. As is true of all thyratrons, once the gas discharge has been started by a positive trigger, the grid has no further control until the discharge is quenched by other means. In modulators employing hydrogen thyratrons, the discharge is quenched when the plate voltage drops nearly to zero after the pulse-forming line has been fully discharged. Hydrogen thyratrons have been employed for developing the large trigger required for the series gap type of switch. Figure 5·25 is the schematic diagram of a hydrogen thyratron modulator giving a 180-kilowatt output pulse 1 microsecond long.

A soft tube switch such as the hydrogen thyratron differs from the other types commonly used with line modulators in that it can conduct current in one direction only. Thus if any situation arises which places a *negative* charge on the pulse-forming network after the pulse is finished, the charge cannot be taken off by the

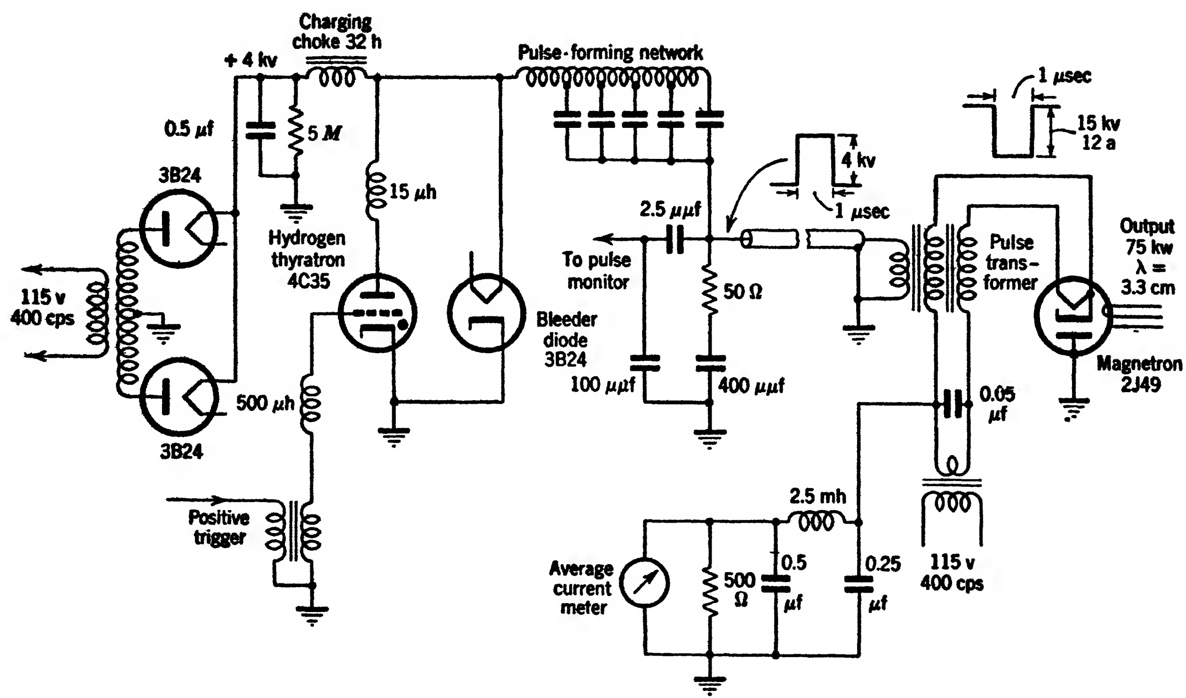


FIG. 5·25 Circuit diagram of a medium power hydrogen thyratron modulator.

thyratron, and it will leak off slowly because of the high impedance reflected through the pulse transformer by the non-oscillating magnetron. The pulse line will then be charged to a higher potential difference in the following interpulse period, and breakdown may occur. A negative charge will be left on the pulse-forming line if the reflected impedance during the pulse falls below the characteristic impedance of the line, as may happen if sparking occurs in the magnetron. The bleeder diode shown in Fig. 5·25 serves to remove this negative charge rapidly.

(*d*) *Pulse-Forming Lines.* The heart of a line-type modulator is the pulse-forming line, which is an artificial transmission line open-circuited at one end. We will show in a qualitative way how an actual transmission line can be employed to form a pulse and will then trace briefly the development of the important types of artificial transmission lines in use at present for high level pulse generation. For the sake of simplicity we will assume the transmission line to be lossless, so that its characteristic impedance Z_0 and delay time τ, are given by (Chapter 2)

$$Z_0 = \sqrt{\frac{L}{C}} \tag{5·17}$$

$$\tau = l\sqrt{LC} \tag{5·18}$$

where L and C are the distributed inductance and capacitance per unit length, and l is the length of the line. Since the delay time is independent of frequency, it is evident that a complex pattern of frequencies, such as that contained in a rectangular pulse, will travel down the line with no disturbance of its phase relations and therefore with no distortion.

From the discussion of transmission lines given in Chapter 2, it can be seen that the incident and reflected signals observed at the open-circuited end of a transmission line have voltages which are in phase but currents which are reversed in phase. Thus a rectangular voltage pulse traveling in one direction on a line has associated with it a current which is equal and opposite to the current associated with an equal voltage pulse of the same sign traveling in the opposite direction. When two such pulses of voltage V and duration corresponding to a length l of the line meet, their voltages add but their currents cancel, so that at the instant of

perfect overlap there is in effect an *electrostatic* charge of voltage $2V$ on a section of line of length l. If a rectangular pulse, of voltage V and width $2l$, is impressed on an open-circuited line, Fig. 5·26 (a), at the instant of half reflection, Fig. 5·26 (c), the pulse width will be l, the pulse voltage $2V$, and the pulse current zero. If the line were suddenly cut at the left edge of the pulse, the electrostatic charge would remain on the line. If, at a later time, the removed section of line were replaced by a resistive load equal to the characteristic impedance of the line, a pulse of amplitude V, current V/Z_0, and duration $2l/v_p$ seconds would appear across the load, v_p being the phase velocity of transmission along the line. At the end of this time the transmission line would be completely discharged, all the energy previously stored in the line as electrostatic charge having been transferred to the load and dissipated therein.

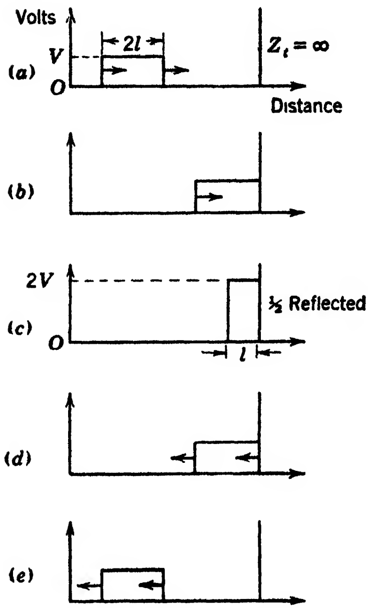


FIG. 5·26 Schematic representation of the reflection of a voltage pulse from the open end of a transmission line. At the instant of half reflection there is in effect an electrostatic charge on the end portion of the line.

This phenomenon is applied in line pulsers. One of the types of switches mentioned above is employed to place across a charged open-circuited transmission line a load having an impedance as nearly as possible equal to the characteristic impedance of the line.

(e) *Artificial Transmission Lines.* Since the pulse duration in seconds is given by $2l/v_p$, and since for ordinary parallel wire and coaxial lines v_p is very nearly equal to $(3 \times 10^8)/\sqrt{K}$ meters per second, K being the dielectric constant of the dielectric material used in the line, it is evident that to form a 1-microsecond pulse an air-dielectric line 150 meters long would have to be used. To

avoid such impractical lengths, artificial transmission lines having *lumped* rather than *distributed* inductance and capacitance have been developed.

Transmission lines can be simulated by low pass filters made up of simple T- or π-sections of the types illustrated in Fig. 5·27; the degree of simulation becomes better the larger the number of sections used. However, it turns out [10] that no matter how many

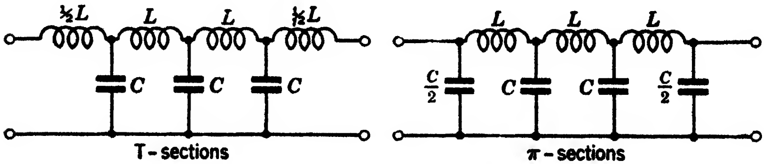


Fig. 5·27 Simple low pass filters which may be used for simulating transmission lines.

sections are employed, the output pulse has a rippled top with the number of ripples corresponding to the number of sections; the initial overshoot is equal to about 18 per cent for a filter containing a large number of sections. These irregularities in the pulse top may be of no consequence if the pulse undergoes further shaping as in a hard tube modulator, but if the pulse is applied either directly or through a pulse transformer to a magnetron, unwanted

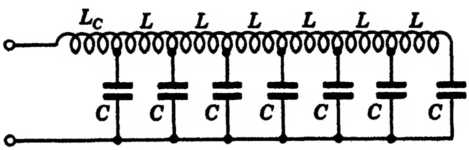


Fig. 5·28 The type E pulse-forming line.

oscillations of the magnetron may result. Considerable attention has therefore been given to the problem of obtaining as good pulse form as possible from a practically realizable artificial line. Detailed analysis of the problem, chiefly by Guillemin, has led to the development of several types of lines. One of the most widely used of these is the type E line illustrated in Fig. 5·28. The design equations for a type E network are

[10] E. A. Guillemin, Radiation Laboratory Report 692, March 1945.

$$Z_0 = \sqrt{\frac{L}{C}}$$

$$\tau = 2n\sqrt{LC}$$

$$\frac{l}{d} = 1.33 \qquad (5 \cdot 19)$$

$$\frac{L_C}{L} = 1.1 \text{ to } 1.2$$

where n is the number of sections, l is the length of coil per section, d is the diameter of the coil, and L and C are the inductance and capacitance per section. The last two equations are empirical. The value of n ranges from 1 to 3 for 0.1-microsecond pulses up to 4 to 7 for 5-microsecond pulses. A five-section line for $\tau = 1$ microsecond, working into a resistive load, gives a pulse which is very nearly flat on top and has a rise time of the order of 0.1 to 0.2 microsecond.

(*f*) *Charging the Pulse-Forming Line.* It is evident from equations 5·19 that

$$nC = \frac{\tau}{2Z_0} \qquad (5 \cdot 20)$$

for a type E line. The capacity nC must be charged to twice the desired pulse voltage in the period between pulses from a suitable power supply. Two types of charging are in general use, one of which employs a d-c supply while the other employs an a-c source.

D-C Charging. In various low power applications such as sweep circuits (cf. Chapter 6) condensers are commonly recharged through series resistances. This practice is not employed with high power modulators, since half the energy supplied from the source is necessarily lost in the resistor. The transfer of charge can be accomplished with very low losses if one employs a series inductance for isolating the line from the supply during the pulse. Qualitatively one may say that the inductance serves as a low impedance to the low frequencies contained in the gradual charging of the line during the interpulse period of hundreds or thousands

of microseconds, but as a high impedance to the high frequencies contained in the short output pulse.

The fundamental d-c charging circuit is represented in Fig. 5·29. Suppose a charge q is taken from the battery; if the losses in the inductance are negligible and if it is assumed that the energy stored in the inductance is the same before and after the charge transfer, one has $q = CV$, where V is the increase in the voltage across the condenser. The energy lost from the battery is qV_s and that gained by the condenser is $\frac{1}{2}CV^2$. Equating these quantities shows that $V = 2V_s$. In actual practice factors in the range 1.85 to 1.95 are obtained because of circuit losses. It is important to note that this voltage step-up is obtained regardless of the size of L, provided there is no net loss or gain of energy in L during the charging period. This situation is realized in two important ways. In *d-c resonant charging*, the value of L is chosen so that the series resonant circuit composed of L and C has a frequency f equal to one-half the pulse repetition frequency f_r:

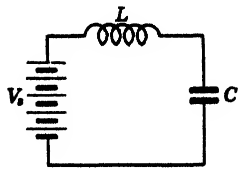


FIG. 5·29 Fundamental circuit for d-c charging of the capacity of a pulse-forming line.

$$f = \frac{1}{2\pi}\sqrt{\frac{1}{LC}} = \frac{1}{2}f_r \tag{5·21}$$

In this case, the voltage across the condenser (or line) and the current through the inductor have the forms shown in Fig. 5·30.

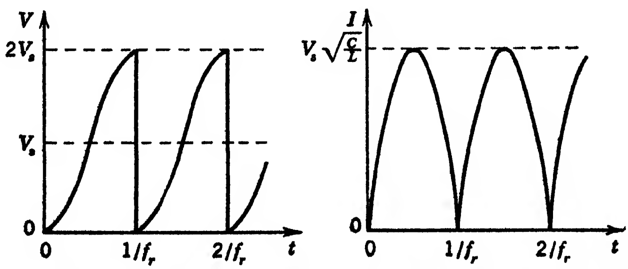


FIG. 5·30 Waveforms in d-c resonant charging of a pulse-forming line.

It is seen that the current is zero at the start and finish of the charging period. In *d-c constant current charging*, on the other

hand, the inductance L is four times or more larger than the resonant charging value, so that the current through it cannot change much, and thus the energy stored in the inductor is practically constant. After the steady state has been reached the voltage and current have the forms shown in Fig. 5·31.

There are certain advantages inherent in each of these types of d-c charging. The resonant method is usually employed with rotating spark gap modulators because the voltage across the line is relatively constant for a short interval in the neighborhood of

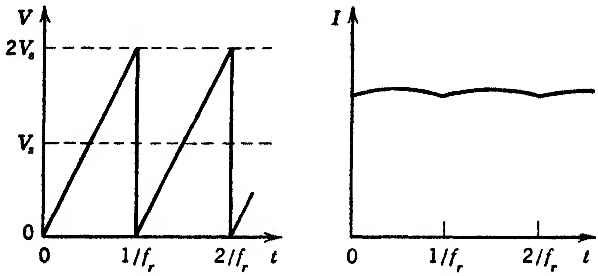


FIG. 5·31 Waveforms in d-c constant current charging of a pulse-forming line.

the discharge time, so that the variations in the pulse amplitude arising from the jitter in the firing of the spark gap are minimized. The constant current method has the advantage that the repetition frequency can be varied over wide limits without changing the value of L, and is somewhat more efficient since the large value of L reduces the magnitudes of the alternating currents flowing in the inductance. Constant current charging is suitable only for use with accurately triggered switches such as the hydrogen thyratron.

A-C CHARGING. It is possible to eliminate the d-c power supply needed in d-c charging, and to charge the pulse line directly from a step-up transformer through a series inductance. This inductance can be built into the transformer as leakage inductance. In a-c charging the recurrence frequency must be an integral multiple or submultiple of the supply frequency, and in most cases is equal to the latter frequency. The necessary a-c is usually obtained from a motor-driven generator, with the rotating member of the spark gap mounted on the same shaft to maintain the proper relation between the supply and recurrence frequencies.

The commonest form of a-c charging is *a-c resonant charging*, in which the value of L (Fig. 5·32) is chosen to resonate with C at the supply frequency. In this case the impressed voltage must be phased so that it is zero at the instant the line is discharged. It can be shown that the condenser recharges in exactly one cycle to a voltage π times the peak impressed voltage, if circuit losses are neglected. Waveforms for a-c resonant charging are shown in Fig. 5·33 for the case of no circuit losses. In actual cases the condenser voltage reaches 90 to 95 per cent of the theoretical value.

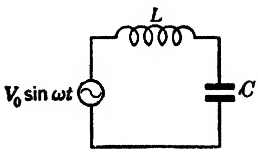


FIG. 5·32 Fundamental circuit for a-c charging of the capacity of a pulse-forming line.

Another case which has some application is one in which the value of L is about one-half that required for the resonant case,

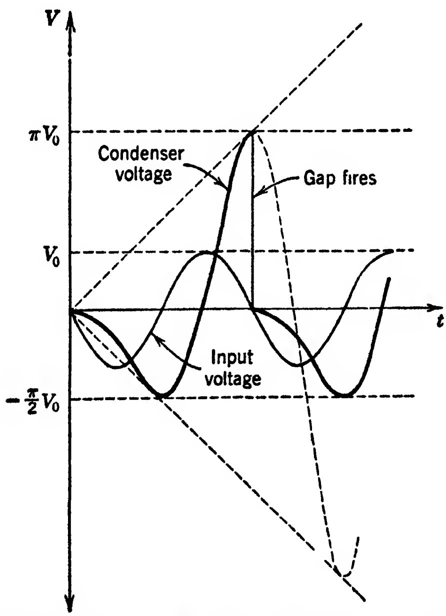


FIG. 5·33 Waveforms for a-c resonant charging of a pulse-forming line.

and the LC circuit has a resonant frequency about 1.4 times the impressed frequency. The step-up ratio is 3.66 instead of π, and

the condenser is discharged at a point close to the maximum of the impressed voltage.

The efficiency realizable with a-c charging is of the same order as that obtained with d-c charging. The a-c case requires considerably simpler equipment in that rectifier tubes and filter circuits are eliminated; on the other hand the d-c method gives a more flexible arrangement so far as variations of repetition frequency and other parameters are concerned.

5·4 VIDEO PULSE DELAY CIRCUITS

Certain applications of pulse circuits involve methods for delaying pulsed signals by more or less accurately controllable periods from less than a microsecond up to thousands of microseconds. For example, in cases where it is desired to view on a cathode ray tube indicator a transient pulse which must itself be used to trigger the indicator sweep, the pulse must be delayed a few microseconds before it is applied to the indicator deflection plates in order to allow time for the sweep to get started.

Delay devices may be classified according to whether they do or do not preserve the input signal waveform. We will consider first three types of circuits which do not preserve waveform and are therefore useful only for delaying trigger pulses. We will then conclude this section with a brief discussion of delay devices which preserve waveform and may therefore be used with video signals as well as with triggers.

Delay Multivibrators

A wide variety of delay circuits which are essentially triggered multivibrators (see Section 10·3) have been devised. In particular, the circuit of Fig. 10·10 can be used for this purpose. In order to develop a delayed positive trigger output, the signal on the first plate of V_2 is differentiated, and the resulting positive pulse is amplified by a tube biased to cutoff, or the differentiation and amplification can be accomplished by a blocking oscillator (cf. page 167). The delay can be varied from a few microseconds to a few milliseconds by proper choice of C and R; in the case of long delays R should be connected to a positive voltage (i.e., +300

volts) instead of to ground in order to reduce jitter. If the delay is to occupy more than about 0.7 of the interpulse period, a circuit similar to that of Fig. 10·11 should be used.

Sawtooth Delay Circuits

A triggered sawtooth generator (Section 5.5) is a device which gives a voltage starting at a fixed value at the time of each trigger

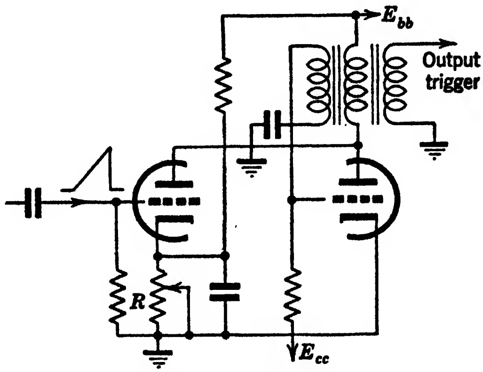


FIG. 5·34 Schematic diagram of a simple sawtooth delay circuit.

and varying linearly with the time for a more or less extended period thereafter. Such a sawtooth voltage can be used as the basis for a variable trigger delay if provision is included for developing an output trigger each time the sawtooth voltage reaches a predetermined value. In the circuit of Fig. 5·34, for example, the bias on the first tube can be varied by means of the cathode resistor *R*. When the sawtooth wave reaches the cutoff potential of the tube, the tube starts to conduct and thus triggers the blocking oscillator. The circuit of Fig. 5·35 illustrates another method of sampling the sawtooth voltage. This method is to be preferred when a precision delay is required. When the sawtooth voltage reaches the diode bias voltage, determined by the setting of the potentiometer *R*, the diode starts to conduct. A high gain pentode (such as 6AC7) amplifies the resulting positive-going signal to form a steeply falling wavefront suitable for accurate triggering of a blocking oscillator. If the highly linear sawtooth produced by a circuit like that shown in Fig. 5·41 is used, and *R* is a precision

potentiometer, the delay time can be made proportional to the potentiometer setting within as little as ±0.1 per cent of the maximum delay time. This type of delay constitutes one of the

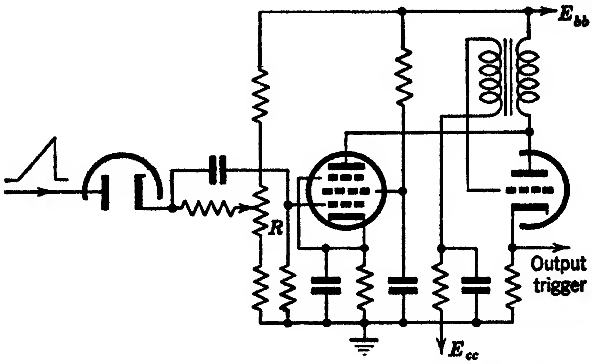


Fig. 5·35 Pick-off diode and output circuit which may be used to obtain a linear delay with the sawtooth voltage developed by the circuit of Fig. 5·41.

most accurately linear delay systems developed for precision range measurements in radar applications.

Phase Shift Delay Circuit

It is evident that the trigger pulses developed from a sine wave by squaring and differentiation come at times dependent on the

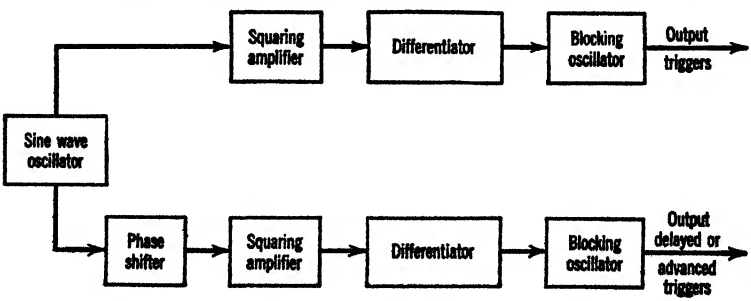


Fig. 5·36 Block diagram of a phase shift delay system.

phase as well as the frequency. A system for producing a train of pulses which can be continuously delayed or advanced in time relative to another train of pulses is outlined in the block diagram

of Fig. 5·36. If considerable care is taken in the design of the squaring and differentiating stages it is possible to obtain triggers

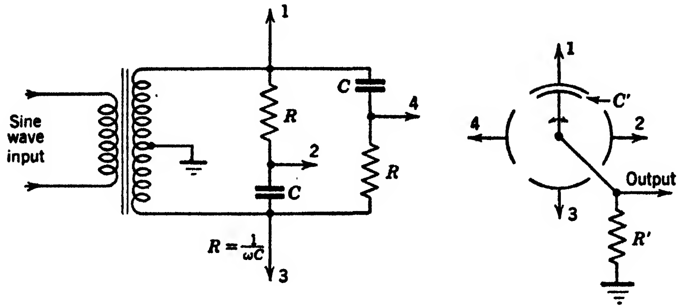


FIG. 5·37 Arrangement for obtaining a continuously variable phase for use in the delay system of Fig. 5·36.

with very little variation in the pulse period. Figure 5·37 illustrates a circuit for producing a sine wave of continuously variable phase.

Electrical Delay Lines

A lossless transmission line (page 175) has a delay time $\tau = l\sqrt{LC}$. With ordinary coaxial lines long lengths are required to give delays of a few microseconds. As shown above, a transmission line may be simulated by lumped inductances and capacitances but, to obtain delays of several microseconds with reasonable freedom from distortion, the line would have to be composed of a large number of sections. Lumped parameter delay lines may, however, be used for delaying trigger pulses for a few microseconds.

Delay lines [11] of relatively large electrical length per unit physical length can be obtained by increasing L and C; for example, a coaxial cable, in which the inner conductor is a tightly wound helix and the outer conductor is a closely fitting braid of insulated wire, can give a delay time of the order of 1 microsecond per foot. Such a line can be used as a variable delay if a small movable pick-up coil is wound around the outer conductor. Delay lines of this general type, having characteristic impedances of about $1K$, are

[11] D. F. Weekes, Radiation Laboratory Report 302, April 1943; H. E. Kallmann, Radiation Laboratory Report 550, June 1944; J. P. Blewett and J. H. Rubel, *Proc. I.R.E.*, **35**, 1580 (1947).

commercially available. Unfortunately, this type of line is characterized by rather high attenuation, and it is difficult to get effective bandwidths in excess of a few megacycles.

Other Delay Systems [12]

Video signals can be delayed for periods up to several milliseconds with a high degree of fidelity by making use of the relatively slow velocity of sound waves in liquid media. For example, a 10- or 15-megacycle oscillator can be modulated by the video signals, and the modulated output used to drive a quartz crystal immersed in mercury. The ultrasonic waves thus excited in the mercury can be received by a second crystal, and the video modulation detected after suitable amplification of the modulated carrier. The velocity of ultrasonic waves in mercury at ordinary temperatures is such that a delay time of approximately 200 microseconds per foot is obtained. This type of video delay system is obviously rather complicated, but it has nevertheless found practical application in a successful system for removing permanent land echoes from the video output of radars (see page 359).

Brief mention may be made of the promising use of so-called storage tubes for obtaining long delays of video signals. It is found that, if the electron beam of a cathode ray tube with a plain glass face is intensity-modulated while the beam is caused to scan the glass surface, a pattern of charges is left on the glass. If at a later time the surface is again scanned with a constant intensity beam, the varying charge pattern is replaced by a uniform pattern; during this process impulses may be obtained from a thin metallic electrode covering the external surface of the tube face which, after amplification, give a reproduction of the intensity modulation initially impressed on the electron beam.

5·5 VIDEO PULSE TIMING

The measurement of the time interval between two pulses is an operation of considerable importance. The determination of the

[12] Detailed discussion of ultrasonic and storage tube delay systems will be found in the *Radiation Laboratory Technical Series*, published by McGraw-Hill Book Co.

distance of an object from a radar set depends on measuring the time interval between the modulator pulse, which simultaneously initiates the transmission of the radar pulse and the radar indicator sweeps, and the echo pulse presented on the indicators. Similarly, a precise method for the determination of the velocity of propagation of ultrasonic waves in a liquid or solid medium depends on video pulse timing. A practical system of radio communication is based on modulation of the time spacing of two or more pulses according to the intelligence to be transmitted.

In making measurements of pulse timing a linear time scale must be established. In general, such a scale is composed of accurately spaced pulses, which may be called *range marks* from their application in radar, and a linear time base for interpolation between range marks. The simplest procedure is to present both the pulse whose time is to be measured and a series of synchronized range marks on an appropriately synchronized cathode ray tube indicator (Chapter 6). If the indicator sweep speed is sufficiently constant, direct interpolation can be accomplished by simple measurement of lengths. Or an auxiliary pulse whose position is controlled by a linear delay circuit (Section 5·3) may be presented on the indicator and used for interpolation. Obviously the design of such a timing system involves many parameters—range mark spacing, sweep speed, and others—which must be carefully chosen to secure the best results. In very precise work attention must be given to equalizing or cancelling delays occurring in amplifiers and other circuits, and pulse shapes must be controlled, particularly with respect to steepness of leading edges, to enable valid position comparisons to be made.

A wide variety of timing schemes has been developed for specialized purposes. For example, some applications, such as the pulse time modulation system of communication (Chapter 12) and automatic radar ranging systems, obviously cannot use any scheme involving manual manipulations.

Range Marks

Range marks of medium stability and accuracy are produced by peaking the output from a shock-excited resonant circuit. Figure 5·38 illustrates such a circuit employing positive feedback to maintain the amplitude of the oscillations. When a negative

unclamping gate (page 307) with steep leading edge is applied to the first triode, the decay of current in the inductance sets up oscillations. The oscillations are rapidly damped out, in spite of the regeneration, when the triode is returned to the conducting region. The frequency of the oscillations can be set equal to the frequency of accurate range marks from a crystal-controlled source (see below) by tuning the inductance. Obviously the range

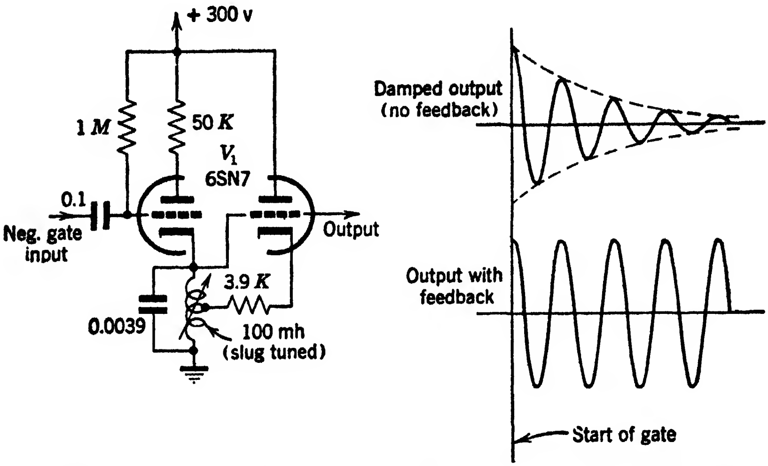


FIG. 5·38 Circuit diagram of a shock-excited oscillator, and representative waveforms.

marks from the shock-excited circuit are tightly synchronized with the trigger initiating the formation of the unclamping gate.

Range marks of considerably greater accuracy but still synchronized with a trigger pulse can be derived from the output of a pulsed crystal-controlled oscillator.[18] A more usual method for obtaining crystal-controlled range marks is to peak the output of a free-running crystal-controlled oscillator; such range marks can be rigidly synchronized with other circuits if the synchronizing trigger is itself derived from the oscillator output by suitable frequency division. The circuit of Fig. 5·39, which is part of the AN/TS-100 A-scope (cf. page 203), illustrates a practical method of accomplishing this result. The trigger repetition frequency is controlled by the blocking oscillator V_{1B}; a small, peaked 80.86-

[18] B. Chance, *Rev. Sci. Instruments*, **17**, 396 (1946).

kilocycle signal from the plate circuit of the crystal oscillator is applied to the blocking oscillator to ensure that it will fire nearly in coincidence with one of the 80.86-kilocycle oscillations. The output of the blocking oscillator is applied as a negative gating pulse to the coincidence tube V_{2A}; the peaked 80.86-kilocycle oscillation which arrives at the grid of V_{2A} while the cathode is

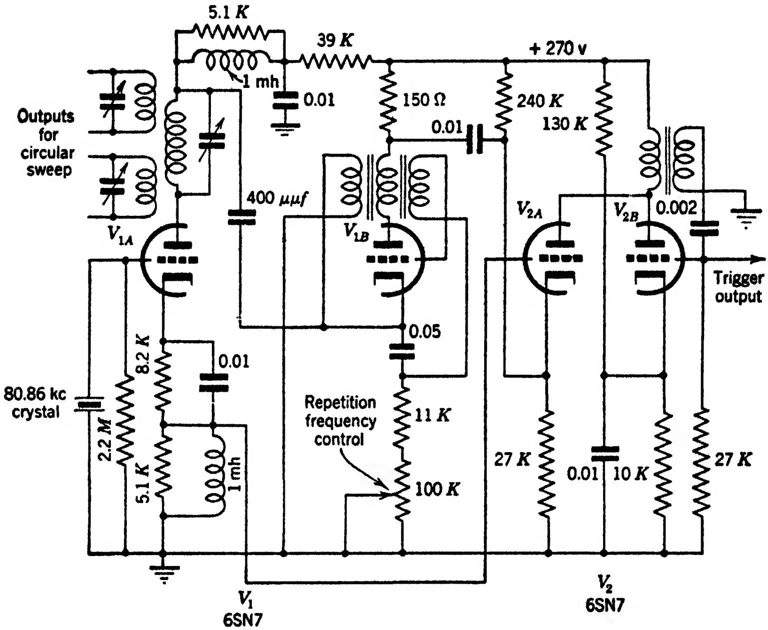


FIG. 5·39 Circuit diagrams of the crystal-controlled oscillator and synchronized trigger generator employed in the AN/TS-100 A-scope.

pulsed down fires the blocking oscillator V_{2B}, the output of which is fed to a cathode follower (not shown) to give a low impedance trigger output. The 5.1K-1mh combinations in Fig. 5·39 constitute strongly damped shock-excited resonant circuits, the inductance and its distributed capacity having a natural period of the order of 1 megacycle. Thus these circuits give a cycle or two of approximately 1-megacycle oscillations once each cycle of the 80.86-kilocycle oscillations.

It should be noted that the frequency of the oscillator V_{1B} is not determined by that of the crystal oscillator, so that there will

usually be a small erratic variation in the spacing of the triggers, although this spacing will always correspond to an integral number of 80.86-kilocycle oscillations. If such jitter in the repetition frequency is objectionable, a counting circuit (page 311) can be employed to reduce the crystal oscillator frequency to the desired low value.

Linear Time Bases

For interpolation between range marks a linear time base is required. The commonest form of time base is a voltage which increases (or decreases) linearly with time. Such a voltage is

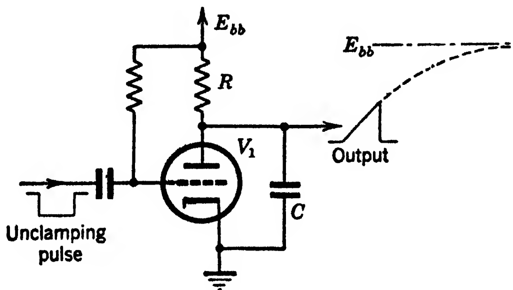


Fig. 5·40 Basic circuit of a simple sawtooth generator.

called a positive-going (or negative-going) sawtooth voltage. For most timing applications the sawtooth voltage must be rigidly synchronizable by means of a trigger pulse, thus differing from the output of the free-running sawtooth generator in the usual cathode ray test oscilloscope.

Many triggered sawtooth generators are of the general type shown in Fig. 5·40. The clamping tube V_1 normally holds the output voltage close to ground. An unclamping pulse, supplied by a triggered multivibrator (page 303), cuts off the tube, and the voltage at the output rises toward E_{bb} along an exponential curve with a time constant equal to RC. The first part of this exponential rise is sufficiently linear for many applications. This simple circuit can be used to produce cathode ray tube sweeps as fast as 10 inches per microsecond.

The linearity of the output of the circuit of Fig. 5·40 can be greatly improved by an ingenious application of feedback. The

circuit is shown in its simplest form in Fig. 5·41. The feedback from the cathode follower V_3 tends to hold constant the voltage drop across, and therefore the current flow through, the charging resistor R. The diode V_2 enables the high potential end of the resistor to rise above E_{bb}; a somewhat lower degree of linearity is obtained if V_2 is replaced by a resistor. The output voltage is given approximately by the expression

$$E_{\text{out}} \approx E_0 + \frac{E_{bb}}{RC} t \tag{5·22}$$

where E_0 is constant and t is the time in microseconds if R is expressed in ohms and C in microfarads. Large deviations from lin-

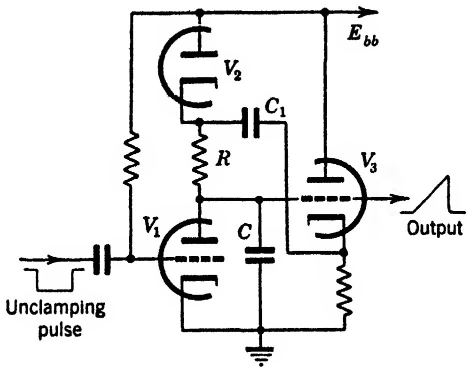


FIG. 5·41 Diagram of a sawtooth generator with regenerative feedback for improving the linearity.

earity will occur if V_3 draws grid current; the voltage at which this takes place can be increased by using a higher plate supply voltage, which can be unregulated, for the cathode follower.

Time interpolations can be carried out by means of a circular sweep on a cathode ray tube equipped with a central deflection electrode (page 197). If such a circular sweep is derived from the output of a crystal-controlled oscillator, as indicated in Fig. 5·39, and if the sine waves in quadrature used to form the sweep are sufficiently free of harmonics to give an accurately circular pattern, a highly precise time base is obtained. Obviously, the pulse interval to be measured must be accurately synchronized with the oscillator producing the sweep, and means have to be provided for

counting the number of revolutions of the sweep between the pulses to be timed. The latter is accomplished, for example, by an auxiliary pulse which can be moved from one pulse to the other around the sweep.

PROBLEMS

5·1 Derive equation 5·1.

5·2 Derive equations 5·3 and 5·4, using the equivalent circuit in Fig. 5·5 (*a*).

5·3 Consider the circuit of Fig. 5·4 with V_1 modified to include a cathode resistor R_K bypassed by a condenser C_K, and with R_L increased to $R_L(1 + g_m R_K)$. Show that the midfrequency gain is

$$A = g_m R_L$$

and the high frequency gain is

$$A_H = \frac{g_m R_L}{1 + 2\pi j f R_L C_S}$$

if $R_K C_K = R_L(1 + g_m R_K)C_S$.

5·4 Derive equations 5·9 and 5·10.

5·5 Derive equations 5·11 and 5·12, using the equivalent circuit in Fig. 5·5 (*b*).

5·6 Derive equations 5·13 and 5·14.

5·7 Suppose that in the circuit of Fig. 5·22 the plate resistor R is large enough so that in the recharging of C between pulses the time constant is practically equal to RC. Derive the steady state value of the difference between the supply voltage E_{bb} and V_C in terms of R, C, ΔV_C and the repetition frequency f_r. Show that if this difference is to be less than $\alpha \Delta V_C$, the repetition period must exceed the value $2.3RC \log (1/\alpha)$. (For the significance of the symbols refer to page 170 of the text.)

5·8 From a consideration of the energy stored in a condenser and the energy of a pulse show that the total charged capacity of a pulse-forming line is given by

$$C = \frac{\tau}{2Z_0}$$

where τ is the pulse width in seconds and Z_0 is the characteristic impedance of the line.

5·9 Derive the expressions for condenser voltage and inductor current during the charging part of the cycle for the case of d-c resonant charging (cf. Figs. 5·29 and 5·30), assuming zero circuit losses. Show that $I_{max} = V_s\sqrt{C/L}$ and $V_{max} = 2V_s$.

5·10 Derive the following expressions for the condenser voltage and inductor current for the charging part of the cycle for the case of d-c constant

current charging (cf. Figs. 5·29 and 5·31), assuming that the circuit losses are zero and that a steady state has been reached:

$$V = V_s\left(\cot \pi \frac{f}{f_r} \sin 2\pi ft + 1 - \cos 2\pi ft\right)$$

$$I = V_s\sqrt{\frac{C}{L}}\left(\cot \pi \frac{f}{f_r} \cos 2\pi ft + \sin 2\pi ft\right)$$

where f is the resonant frequency of the L–C combination. Show that

$$\frac{I_0}{I_{\max}} = \cos \pi \frac{f}{f_r}$$

where I_0 is the current immediately after each discharge of the condenser, and estimate how much larger L must be than for the resonant case (Problem 5·9) to have the current constant within ±5 per cent. Show that $V = 2V_s$ at the end of each charging cycle regardless of the value of L.

5·11 Show that the condenser voltage during the charging period in a-c resonant charging (cf. Figs. 5·32 and 5·33) is given by the expression

$$V = \frac{V_0}{2}(\omega t \cos \omega t - \sin \omega t)$$

when the input voltage is

$$V = -V_0 \sin \omega t$$

provided circuit losses can be neglected.

5·12 Show that the output frequency of the circuit of Fig. 5·38 without feedback is given very closely by $1/2\pi\sqrt{LC}$ if the circuit losses are low, and derive the expression for the time constant of the decay envelope. Assume that the low circuit losses are due entirely to the resistance of the inductor.

5·13 Show how the circuit of Fig. 5·40 can be modified to produce a negative-going sawtooth voltage.

CHAPTER 6

CATHODE RAY TUBE INDICATORS

The cathode ray tube (CRT) indicator is an instrument of central importance in modern electronics. It is a *sine qua non* to radar, which could certainly never have developed beyond an exceedingly clumsy device had it not been possible to present rapidly a mass of information in clearly readable form on the face of a CRT indicator. Nearly all the non-radar applications of microwave and electronic techniques mentioned elsewhere in this book involve the display of useful information by means of a CRT.

A CRT indicator is essentially an extremely high speed writing device. A readily visible spot of light is produced by the bombardment of a fluorescent screen with a focused stream of rapidly moving electrons. The great value of the CRT is that this stream of electrons can be rapidly deflected from its normal course by the application of electric or magnetic fields of moderate intensity; because of the very small mass of the electron, sizable deflections can be produced in very short time intervals and with the expenditure of very small amounts of energy. The light spot can be made to travel across the tube face at rates as high as several hundred inches per microsecond, and the driving force required (in the case of deflection by an electrostatic field) is almost entirely that required to charge the capacity of the deflection plates and associated wiring independently of the presence of the electron beam.

6·1 CATHODE RAY TUBES

The important types of cathode ray tubes may be classified according to two broad schemes: the type of focusing and deflection used, and the type of fluorescent material or *phosphor* employed.

Electrostatic Tubes

The tubes used in ordinary oscilloscopes and certain radar indicators employ electrostatic focusing and deflection of the electron beam. A typical design is shown in Fig. 6·1. The cathode is an

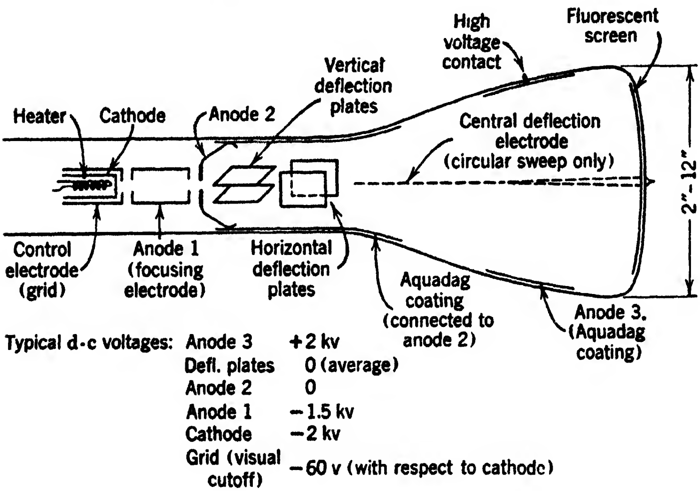


Fig. 6·1 Schematic representation of a typical electrostatic cathode ray tube.

oxide-coated cylinder heated by a filament, and is normally held at a potential of 1 to 2 kilovolts below ground. The flow of electrons from the cathode is controlled by the potential of a grid cylinder, and the beam is brought to a focus on the fluorescent screen by the electrostatic field [1] generated by the focusing electrode or first anode. This anode in the case illustrated is at a potential approximately 500 volts above the cathode; a front

[1] For a discussion of electron optics see I. G. Maloff and D. W. Epstein, *Electron Optics in Television*, McGraw-Hill Book Co., 1938.

panel control of this voltage is required to allow careful adjustment. After being focused, the electrons are accelerated in the field between the first and second anodes. The latter anode, which thus serves as a *predeflection* accelerating electrode, is normally at ground potential. The accelerated electrons then pass through the deflection plate system. After deflection the electrons may be further accelerated by a *postdeflection* electrode or intensifier. This third anode is a coating of Aquadag (colloidal graphite) on the inside of the tube, and in the case illustrated is held at a potential of about 2 kilovolts above ground. The advantage of a postdeflection accelerating electrode is that the additional acceleration is obtained with only a small decrease in the deflection sensitivity. The tube illustrated has a deflection sensitivity, with the third anode at ground potential, of approximately 0.4 millimeter deflection per volt; this figure is decreased to about 0.3 millimeter per volt on raising the third anode to +2 kilovolts. With these relatively high accelerating potentials, the grid potential must be lowered to 60 volts below the cathode potential to achieve visual extinction of the beam (visual cutoff).

The brightness of the spot and the sharpness of focus improve as the accelerating potentials increase. This is important in many applications where very high writing speeds are employed and where the maximum resolution of signals on the tube face is desired. The use of high accelerating potentials has been carried to an advanced point in the development of types such as the 5R series of tubes.[2] These tubes have a series of three postdeflection accelerating electrodes, to the last of which a potential as high as 20 kilovolts may be applied, the deflection sensitivity being of the order of 0.15 millimeter per volt. With such tubes extremely bright traces can be obtained with very high writing speeds.

In the application of electrostatic tubes, sweep voltages are customarily applied to the horizontal deflection plates, and signal voltages to the vertical plates. Usually the sweep voltage is a linear function of time in order to give a linear time base. The application of the sweep and signal voltages interferes less with the focus of the beam if they are applied in a push-pull manner; that is, if the potential of one plate is decreased at each instant by the amount the potential of the other plate is increased. The use of

[2] See page 200 for a description of CRT-type designation symbols.

push-pull deflection is more important in the sweep voltage since it generally is considerably larger than the signal voltage.

The useful sweep length with a given tube can be increased by using a circular rather than a linear sweep. A circular sweep is produced by putting a sine voltage on one set of plates and a voltage of the same frequency differing in phase by $\pi/2$ on the other set. The sweep speed in this case can be very accurately controlled by deriving the two voltages from a crystal-controlled oscillator; for this reason, this type of sweep is frequently employed in the precise measurement of time intervals (see page 191). Deflection modulation of the beam is accomplished by impressing signals on a central deflection electrode projecting through the tube face (shown dotted in Fig. 6·1). With this type of electrode the deflection sensitivities are considerably lower than those obtained with the deflection plates.

Magnetic Tubes

Cathode ray tubes in which the electron beam is focused and deflected by *magnetic* fields find wide application, particularly as

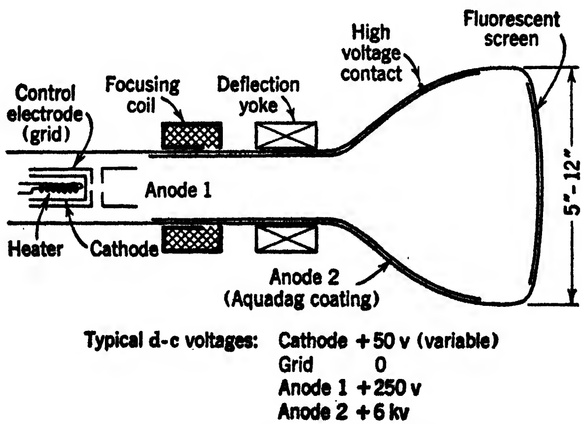


FIG. 6·2 Schematic representation of a typical cathode ray tube employing magnetic deflection.

radar indicators. Figure 6·2 illustrates schematically a magnetic tube. The electron "gun" is similar to that used in electrostatic tubes. After the grid there is usually a first anode (sometimes

called the screen grid) which serves to make the cutoff bias independent of the total accelerating voltage and which is also useful in controlling the beam current. This anode is held at a potential approximately 200 volts above the cathode during normal operation. The grid bias then required for visual cutoff is of the order of −45 volts. The beam is focused in the axially symmetrical field of the focus coil, which is usually a multilayer coil enclosed in a soft iron shell in the inner wall of which there is a gap; the leakage flux across this gap provides the focusing field. The current through the focusing coil is adjusted by a control on the front panel.

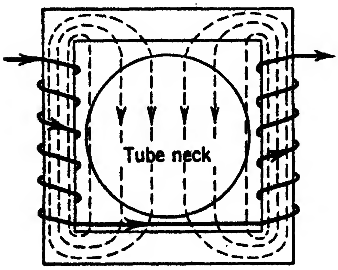


Fig. 6·3 Simple type of deflection yoke for a magnetic cathode ray tube.

The focused beam is deflected by means of the deflection coil or yoke, which may take many forms depending on such factors as the purpose for which the indicator is intended. The simplest form is shown in Fig. 6·3. Two windings on a square iron core are arranged with their fluxes opposed. There is thus produced a leakage field as indicated across the neck of the tube, which must be uniform to avoid distortion of the beam. Sometimes multiple coils are used to permit beam centering or off-centering and the use of push-pull sweeps. With the yoke oriented as shown in the figure, the beam is deflected in the horizontal direction. The amount of deflection will increase at a uniform rate, thus producing a linear time base on the tube face, if the magnetic field produced by the yoke increases linearly with time. This requires that the *current* through the deflection coils increase linearly; that is, a sawtooth current waveform is required.

The direction of the sweep produced by the yoke illustrated in Fig. 6·3 can be changed by mechanical rotation of the yoke. This constitutes one of the most satisfactory ways of producing the rotating linear sweep used in Plan Position Indicators (page 206).

The electrons in the beam are accelerated by means of a high positive potential applied to the second anode, which consists of an Aquadag coating on the inside of the tube extending from well down in the neck nearly to the face of the tube. The second

anode potential is usually of the order of 6 kilovolts above the cathode.

Cathode Ray Tube Screens

Our understanding of the fluorescence of materials used in cathode ray tube screens is very slight, in spite of the considerable amount of work which has been done on the subject. It is known that the behavior of a fluorescent layer, or phosphor, is greatly dependent on the nature and amount of the impurities present in it. Thus extremely pure zinc orthosilicate shows very little fluorescence on bombardment by electrons but, when "activated" by traces of silver, it fluoresces with a green light.

As commonly used in the present connection, the terms fluorescence and phosphorescence are chiefly distinguished by the rate of decay, or *persistence*, of the light produced. The decay of fluorescence requires times of the order of milliseconds, whereas phosphorescence may persist with visible intensity for a period of many seconds.

Several types of phosphor have been developed for various applications. Any of these types may be employed with either electrostatic or magnetic tubes. A property of phosphors which has great importance in radar and other applications is the persistence, which may be quantitatively expressed in terms of a decay law. The light from some phosphors decays according to the expression

$$I = I_0 t^{-s} \tag{6·1}$$

where I is the light intensity at time t and I_0 that at time $t = 0$, and s is an empirically determined constant. Other phosphors follow an exponential decay law:

$$I = I_0 e^{-t/\tau} \tag{6·2}$$

where τ is the decay time constant. The persistence of a screen determines whether it is suitable for applications in which rapidly changing signals occur in a given region of the tube face, or whether it may be applied in cases where there is a more or less extended period between successive excitations of a given region. An example of the former type of application is an A-scope (page 203) used with a radar having a scanning antenna, so that a constantly

changing signal pattern is displayed on a horizontal sweep. The latter type is illustrated by the PPI (page 206) of a slowly scanning radar; here the linear sweep rotates in synchronism with the antenna, and there is an interval of several seconds between successive passes of the sweep over a particular spot on the tube face. In such cases it is necessary to have an excited spot continue to emit visible light for a large fraction of the period of rotation of the sweep.

Long persistence may be achieved by the use of a *cascade* screen consisting of two separate layers. The layer closer to the electron gun fluoresces under electron bombardment, and the other one phosphoresces with long persistence when excited by the light from the first layer.

It is found that some phosphors are excited to practically their saturation intensity (for a given accelerating voltage and beam current) by a very short period of excitation, of the order of microseconds in duration. Other phosphors, in contrast, require more or less extended periods of excitation to reach saturation. Such phosphors usually show an integration phenomenon known as *build-up*. If the electron beam is pulsed on for much less time than required to reach saturation intensity, it is found that the ratio of the intensity resulting from excitation by n pulses to that resulting from one pulse is greater than n. This integration effect results in a complicated dependence of signal visibility, particularly in the presence of random "noise" pulses, on such parameters as pulse width and pulse repetition frequency. Build-up may be expressed quantitatively in terms of the *build-up ratio*, which is the ratio of the intensity remaining at a standard time (usually 1 second) after n standard excitations repeated at a fixed frequency to the intensity the same time after one excitation, and is represented by the symbol $G_{n:1}$.

Table 6·1 gives some of the properties of important types of screens. The P1 and P7 screens are much the most important in radar applications. Several other screens are available, such as the P4 screen, which gives a white light and is used in television picture tubes.

RMA Designation of Cathode Ray Tubes

The Radio Manufacturers' Association designation of a CRT gives first the nominal screen diameter in inches, then a letter in-

TABLE 6·1 CATHODE RAY TUBE SCREENS *

Phosphor Designation	Type	Composition	Color of Light Emitted	Decay	Build-up	Remarks
P1	Single layer	Zn_2SiO_4:Mn	Green	$I = I_0e^{-t/\tau}$ $\tau = 18$ msec	No useful build-up	Medium persistence; objectionable flicker below 20 cps; photography on moving film up to 40 cps. Used in test and A-scopes.
P2	Single layer	ZnS:CdS:Mn:Cu	Yellow-green	$I = I_0t^{-S}$ $S = 0.7$	$G_{5:1} = 1.9$ $G_{100:1} = 3.6$	Long persistence; useful in region of 1 scan per sec.
P5	Single layer	$CaWO_4$	Blue	Most rapid of any production phosphor	No useful build-up	Very short persistence; photography on moving film up to 60 kcps.
P7	Cascade	ZnS:Ag (fluorescent) ZnS:CdS:Cu (phosphorescent)	Blue Yellow-green	$I = I_0t^{-S}$ $S = 0.5$	$G_{5:1} = 10$ $G_{100:1} = 40$	Long persistence; used in PPI's and B-scopes in region of 2–60 scans per min.
P10	Single layer	KCl	Magenta darkening of white screen	Variable from that of P7 to several days depending on "burn-in"		So-called "dark trace" tube; used in one form of projection PPI.

* Most of the material in Table 6·1 is taken from Radiation Laboratory Report S–48, by A. B. White, May 1945. A detailed discussion of the P7 cascade screen is given by W. B. Nottingham, Radiation Laboratory Report VI–4S.

dicating merely the order of development or registration of that particular type of tube, and finally the phosphor designation. Thus a 3BP1 tube has a 3-inch screen and uses the P1 phosphor.

Control Grid Characteristics

A typical family of control grid voltage versus beam current curves is given in Fig. 6·4. Voltages are expressed relative to the cathode voltage. Aside from scale changes, these curves are simi-

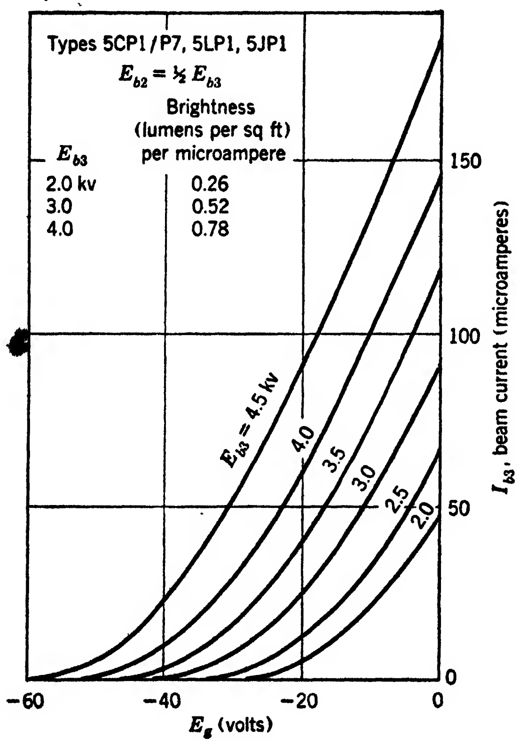


FIG. 6·4 Typical grid voltage-beam current characteristics for cathode ray tube types 5CP1/P7, 5LP1, and 5JP1.

lar to those obtained with ordinary receiving-type tubes. As indicated by the table included in the figure, the screen brightness is proportional to the beam current and, in the region covered by the table, increases linearly with the total accelerating voltage.

6·2 TYPES OF INDICATORS

Two types of modulation can be applied to the electron beams in electrostatic or magnetic tubes: deflection and intensity modulation. The means for applying deflection modulation have already been described. Intensity modulation is in most cases accomplished by varying the grid or cathode potential, or both. Relatively long period modulation such as gating is frequently brought about by changing the screen grid potential. Two signal trains for intensity modulation can be conveniently mixed by applying one to the grid and the other, in opposite polarity, to the cathode. Similarly, in deflection modulation, one set of signals can be applied to one of the vertical deflection plates and the other to the other plate.

The following paragraphs give brief descriptions of some important types of CRT indicators. These types are for the most part selected from the wide variety of radar indicators which have been developed for displaying in various ways information about the position of targets. None of these types, except for the A-scope, has as yet found extensive non-radar application, although it is quite probable that some of them will.

A-Scope

An A-scope is an indicator employing an electrostatic tube with a linear sweep which is initiated by a trigger. It thus differs from an ordinary oscilloscope mainly in that the sweep can be tightly synchronized with an external system. This type of indicator is sometimes called a *synchroscope* and is indispensable in work with pulse circuits. In radar practice, A-scopes are extensively used for such purposes as tuning radar systems (since changes in the amplitude of a signal displayed as a deflection modulation are more readily observed than changes in intensity) and range measurements. The position in range of a signal displayed on an A-scope can be accurately determined by comparison with a movable range mark or notch generated by a calibrated delay circuit (see Chapter 5). Several modifications of the simple A-scope have been developed for this purpose. Thus, the sweep may be rapid, say ⅕ inch per microsecond, and delayed a variable but accurately known time after the instant the radar pulse is transmitted.

The delay may be controlled by a handwheel and the position of a marker which is brought into coincidence with the signal read from a set of dials directly in yards or miles. In some A-scopes, provision is made for expanding a small portion of the sweep

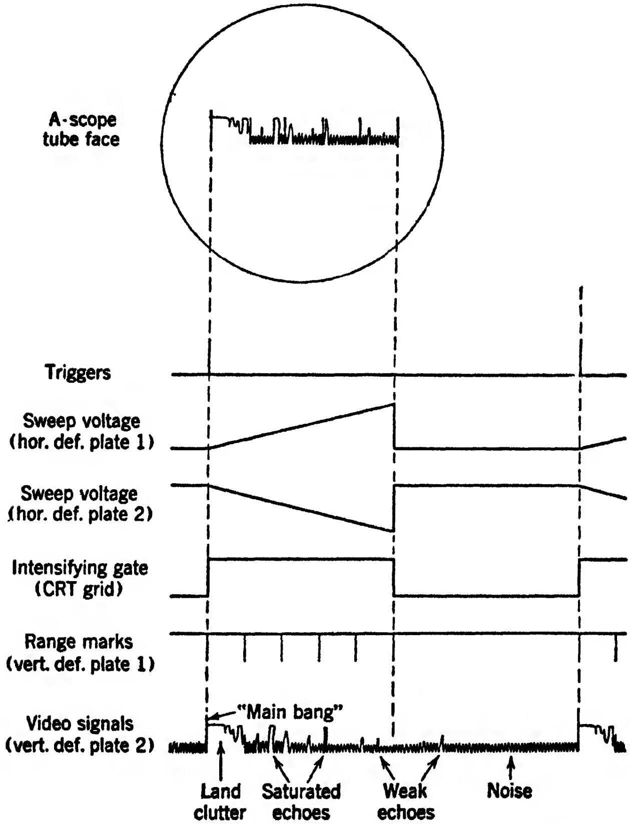


FIG. 6·5 Representation of the appearance of a type A presentation, showing important waveforms.

including the signal, the position of the expanded portion of the sweep again being controlled by the operator and its position being read from appropriate dials. In nearly all cases, A-scopes employ screens of medium persistence, such as the P1 screen.

Figure 6·5 represents the appearance of an A-scope on which are presented the signals from a radar. The waveforms shown in

the figure illustrate the time relations of the signals fed to the various CRT electrodes. The generation of these waveforms is discussed in Section 6·3. The radiated radar pulse is frequently called the "main bang"; the video signal corresponding to the main bang usually has an amplitude exceeding the normal video limit level of the system because of direct feed-through by ground currents from the enormous video pulse applied to the transmitter tube (cf. Section 5·3).

The synchroscope provides the best type of display for many time measurements on either single or repeated transient phenomena and is thus of great importance in scientific work. It is important to note in this connection that the development of the synchroscope for extremely short interval timing has been carried to an advanced state. Lee [3] has described an oscilloscope of the Du Four type, in which the electron beam impinges directly on a photographic film inside the evacuated tube, with which one can actually obtain an oscillogram showing a few cycles of a 3000-megacycle sine wave! More recently, Winter [4] has described a synchroscope with sweep speeds up to 100 inches per microsecond which employs a fluorescent screen giving sufficient intensity so that single transients involving writing speeds up to 300 inches per microsecond can be recorded on an external camera. With this device a 100-megacycle sine wave can be readily recorded, and the time interval between two points of a single transient can be stated with a precision of about $\pm 10^{-9}$ second. Needless to say, very considerable refinement over ordinary circuit methods is required to achieve operation of this sort.

J-Scope

The J-scope is essentially an A-scope with a circular sweep, produced as described on page 197. Very precise ranging, to 20 yards or so, may be accomplished by using a crystal-controlled 2000-yard sweep on the J-scope; the radar trigger is produced by

[3] G. M. Lee, Master's Thesis, M.I.T., June 1944; *Proc. I.R.E.*, **34**, 121W (1946).

[4] D. F. Winter, Radiation Laboratory Report 1001, April 1946. V. L. Fitch and E. W. Titterton, *Rev. Sci. Instruments*, 18, 821 (1947), have given a complete description of a synchroscope having a much lower maximum writing speed, but including several useful accessories.

counting down (cf. page 311) the crystal oscillator output to a suitable frequency, so that the scope sweep and radar are rigidly synchronized. The range to a target is determined by counting the number of revolutions and the fraction of one revolution before the echo appears on the indicator. As in A-scopes, medium persistence screens are employed.

PPI

The Plan Position Indicator, or PPI, is of great importance in radar because it gives a nearly true map picture of the surroundings. The general usefulness of the PPI is indicated by the fact that on naval ships numerous remote PPI's are located at various points from the bridge down. These remote indicators are arranged so that they can be switched by a single knob to any one of the various search radars on board.

PPI's have a linear sweep from the center of the tube to the edge, which rotates around the tube face in synchronism with the radar antenna. Almost without exception they employ magnetic tubes. The grid bias is adjusted so that the sweep is barely visible in the absence of signals or noise. Signals (and noise) are applied to either the grid or the cathode in the proper polarity to intensify the beam. The usual signal amplitude adjustment is such as to have the noise peaks give a slight general brightening of the trace; it is necessary to have the noise large enough to brighten the trace appreciably in order to avoid losing weak signals.

If the beam intensity is increased too much the focus is destroyed; the spot is said to "bloom." It is therefore necessary to include a limiting stage in the video amplifier preceding the PPI to prevent strong signals from causing blooming.

In addition to the intensity modulation produced by the signals, the electron beam is blanked out during the part of the radar repetition interval when the indicator is not being used. This is accomplished by applying a suitable gating signal to an electrode such as the first anode. In addition, range marks may be displayed on the screen by feeding short pulses to either the grid or the cathode (whichever is not used for the signals), and in some radars azimuth marks are also included at intervals such as every 10 degrees of antenna rotation.

Since the sweep is in a given position for only a short time, it would be impossible to "read" a PPI if the light produced by a signal did not persist for some time. For this reason, a screen such as the P7, having long persistence, is always used in this type of indicator.

PPI's employ tubes varying in diameter from 5 to 12 inches. There is a general preference for 12-inch tubes in cases where

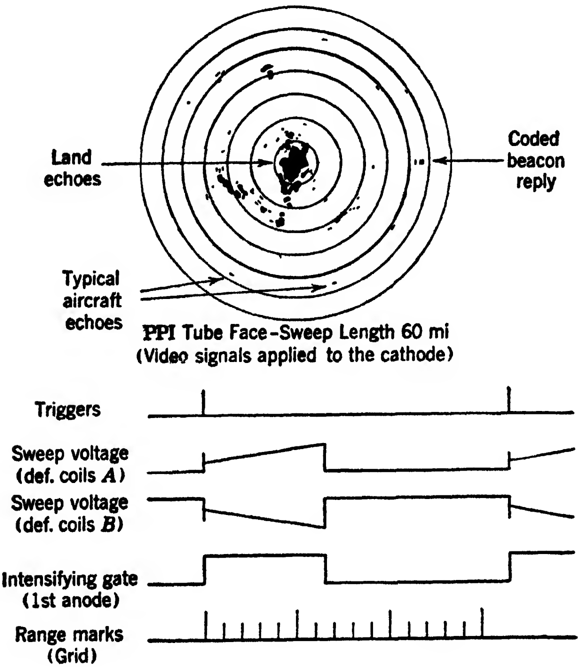


FIG. 6·6 Representation of the PPI of a land-based radar. The tube face is shown as it appears in a photograph of a complete rotation of the sweep; i.e., the instantaneous position of the sweep and the decay of the signals are not shown. The time relations of the signals applied to the tube are indicated by the waveforms given.

space and weight limitations do not enter. The spot size increases with the tube size, so that there is little or no gain in resolution in going to large tubes; they are preferred because they are somewhat easier to read, and plots and other marks may be more easily made by the operator on the tube face.

Figure 6·6 gives a representation of a PPI display (for convenience signals are shown as dark spots on a white background),

together with significant waveforms to illustrate the timing of the various voltages and currents impressed on the indicator.

Since the PPI presents range radially and azimuth as an angle, it gives a true map picture of objects located in the horizontal plane of the radar. The range and azimuth of all objects within view are, of course, presented correctly, but no distinction is made between *slant* range and *horizontal* range. Thus, if an aircraft at a high altitude is at the same bearing as a hill, and is at a *slant* range equal to the range of the hill, its echo will coincide with that of the hill even though it is flying over a point appreciably closer to the radar. The approximately true map nature of the PPI picture makes it particularly valuable in ship navigation. Optical and electronic systems have been devised which permit superimposing a portion of a map or chart on the PPI tube face to aid in recognizing the terrain features shown by the indicator.

B-Scope

A B-scope gives a rectangular plot of ranges *versus* azimuth, range usually being represented in the vertical direction on the tube face and azimuth in the horizontal direction. The picture presented is thus highly distorted, but it is nevertheless very useful in enabling operators to read off the ranges and bearings of targets. The B-scope is usually designed to display a small range interval of 10 or 20 miles and an approximately 90° sector of azimuth. The particular region shown on the indicator may be selected by the operator; suitable calibrated dials incorporated in the controls, together with electronic range and angle markers, facilitate the determination of the coordinates of a signal.

Type B indicators usually employ magnetic deflection of the electronic beam. The range sweep is produced much as in a PPI. This sweep is moved relatively slowly across the tube face at a rate determined by the rate of rotation of the radar antenna and the size of the sector covered; a circuit for supplying the azimuth sweep current is described on page 216. The electron beam is blanked out during the fraction of each repetition interval when the range sweep is not operative, and the fraction of each antenna rotation interval when the azimuth sweep is not operative.

As in the case of the PPI, B-scans employ long-persistent screens in sizes up to 12 inches.

Precision PPI

The Precision PPI, or P^3I, is a combination of a PPI and a B-scope which is useful for precise range and angle measurements in conjunction with a search radar which supplies sufficiently accurate angle information from the antenna mount. The PPI is used for selecting targets whose coordinates are to be measured. A range "strobe" or marker indicates on the PPI, in the form of a slightly brightened band, the range segment covered by the B-scope, and a mechanical angle marker indicates the center of the azimuth sector covered. By means of handwheels these markers are moved to include the desired target, which will then appear on the B-scope when the antenna reaches the proper angular position. On the B-scope there is a fine electronic range strobe and an angle marker, the positions of which are controlled by the same handwheels. The range strobe remains on the tube face after the azimuth sweep is completed, and it and the angle marker can be moved to intersect precisely at the persistent target trace. The handwheels are coupled with mechanical revolution counters or dials which show the target range in yards and bearing in degrees after the above operation has been completed. Circuits are included which move the intersection of the range and angle marks and the signal to the center of the B-scope at the next rotation of the antenna.

Range Height Indicator

As its name implies, this indicator, abbreviated RHI, displays the two coordinates slant range and altitude. The RHI is employed with height-finding radars of the "beavertail" type (page 349) in which the antenna, which has a narrow beam in elevation, is rapidly scanned in elevation. In effect, a linear range sweep is rotated about a point at the lower left-hand side of the tube, in synchronism with the elevation scanning of the antenna, from slightly below the horizon to 30 degrees or so above. In one type of RHI, altitudes are read from an appropriate scale over the tube face. As a rule, electronic range marks are shown on the tube, and intensity modulation is employed for both the range marks and the video signals.

In one form of RHI linear horizontal and vertical sweeps are used, both of which occur at the radar repetition frequency. If

the speed of the horizontal sweep is ρ inches per microsecond, the speed of the vertical sweep is approximately $\rho(\sin \phi / \sin \phi_{max})$ inches per microsecond where ϕ is the antenna elevation. It is evident that the horizontal deflection to an echo measures the slant range, and the vertical deflection the altitude. If a long range interval is covered, the graduation lines of constant altitude have to be curved to allow for the earth's curvature.

6·3 INDICATOR CIRCUITS

A complete discussion of the electronic circuits employed in indicators would require far more space than is available here. A selection of typical circuits will be described in sufficient detail to illustrate the problems which arise and some of the methods developed for solving them.

Gating Pulses

Gating or switching pulses are frequently employed in indicators in connection with electronic switches (clamps; see page 307) and for blanking or intensifying the electron beams in cathode ray tubes. These gates are generated by some form of triggered multivibrator, such as that described on page 303.

Range and Angle Marks

Electronic marks to show range and angle intervals are used in many indicators. Circuits for the generation of range marks are described in Section 5·5. Angle marks can be generated, in slowly scanning radars, by means of a photocell-slotted disk arrangement. The disk is driven through a servo link (Chapter 9), usually the same link as used to drive the PPI's, in synchronism with the antenna. The angle marks consist of pulses two or three times as long as the radar repetition interval, or of oscillations (say at 100 or 200 kilocycles) which have this envelope and are not synchronized with the radar pulse rate.

Range Sweeps, Electrostatic

As mentioned above, range sweeps, or linear time base sweeps, in the case of electrostatic deflection are obtained by imposing a

sawtooth voltage between a pair of deflection plates. The generation of such voltage waves increasing linearly with time is discussed in Chapter 5. In cases where a push-pull sweep is used, an inverter stage feeds the sweep signal to one of the deflection plates while the other plate receives the sawtooth direct.

Linear sweeps for extremely short interval timing[5] involve circuits somewhat different from those described in Chapter 5. A very fast-acting switch tube is required, and relatively large sweep power is necessary to charge the various lumped and distributed capacities in the sweep circuit. Such requirements have been met by adopting many of the techniques employed in high level pulse generators.

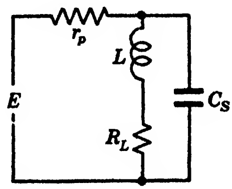


FIG. 6·7 The approximate equivalent circuit for the driver tube and deflection coil used in obtaining a linear sweep in a magnetic cathode ray tube.

Range Sweeps, Magnetic

If the electron beam in a CRT is to be deflected at a constant rate by means of a magnetic field, the field intensity must also change at a constant rate. This means that the *current* through the deflection coils must change linearly with time. The problem of obtaining such sawtooth current waves is considerably more complicated than that of obtaining sawtooth voltage waves in the electrostatic case, where the impedance to be driven is very large.

The approximate equivalent circuit for the sweep driver tube and the deflection coils is shown in Fig. 6·7. The coils have inductance L and resistance R_L, and are shunted by the wiring and coil capacity C_S. The plate resistance r_p of the driver stage is in series with the L, R_L, and C_S combination. If we neglect the capacity, we have

$$E = L\frac{di}{dt} + (r_p + R_L)i \qquad (6\cdot3)$$

For a linearly increasing current, $i = kt$, where t is time:

$$E = kL + k(r_p + R_L)t \qquad (6\cdot4)$$

[5] D. F. Winter, Radiation Laboratory Report 1001, April 1946.

Thus the applied voltage should have a trapezoidal form. Such a trapezoidal voltage is approximated by the output of the circuit shown in Fig. 6·8. A negative gate of the same duration as the desired sweep is applied to the normally conducting first tube. A rapid voltage rise at the plate of this tube results because the capacity C is a very low impedance to short transients; the magnitude of this initial rise is approximately $E_{bb}[R_2/(R_1 + R_2)]$. The condenser voltage then starts to rise toward E_{bb} with the time

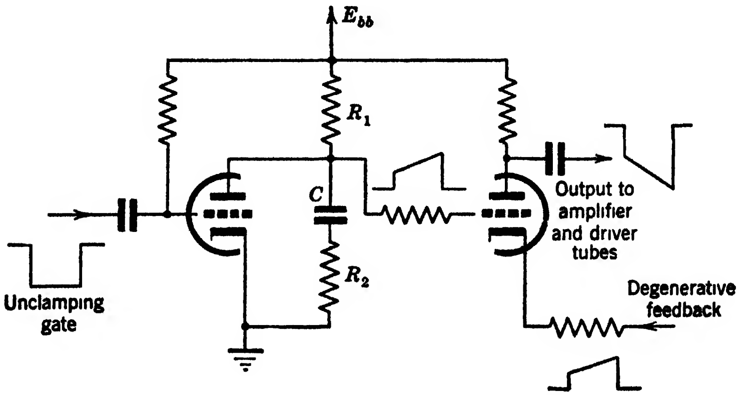


FIG. 6·8 Schematic diagram of a simple circuit for obtaining the trapezoidal voltage for driving a deflection yoke.

constant $(R_1 + R_2)C$, the first part of this exponential rise being sufficiently linear for most purposes. The trapezoidal voltage is then amplified and finally applied to the grid of a driver tube having the deflection coil in its plate circuit. To improve the linearity of the sweep amplifier stages, degenerative feedback from the driver is applied to the cathode of the first amplifier stage. In most cases it is necessary to place a diode d-c restorer (page 307) in the grid circuit of the driver stage to re-establish the proper zero point for the beam between sweeps.

Rotation of Linear Sweeps

(*a*) The simplest method of obtaining a rotating linear sweep with a magnetic tube is to rotate the deflection yoke, the sweep signal being fed to the deflection coils by means of slip rings. In radar applications, this method requires some sort of angle data

transmission system between the antenna and the deflection yoke. A very simple arrangement is a 1-speed synchro link (see page 275); for example a synchro generator might be geared to the antenna shaft and a synchro motor used to drive the deflection yoke. In order to have a one-to-one correlation between antenna and yoke positions, the synchro units must rotate through one revolution for each antenna revolution. Such a system will be subject to errors of 2° or more.

Where greater accuracy is needed, which is usually the case, synchros turning at ten to thirty-six times the antenna rotation speed are used. The error is reduced as the synchro speed is increased; thus in a 10-speed system the errors should not exceed a few tenths of a degree. Because a 10-speed synchro system can lock in equally well at any one of 10 positions, additional provision must be made to secure a one-to-one correlation between antenna and indicator positions. A simple mechanical switching system to accomplish this is shown in Fig. 6·9. The 1-speed shafts at the antenna and the indicator are equipped with cams which actuate microswitches for 18° and 14° respectively, the cams being centered at north. If the indicator is properly synchronized the two switches are never closed simultaneously, but if they are more than about 16° out of step, which will only happen when they are locked in one of the nine wrong relative positions, the relay will be energized when the PPI reaches its north position, and S_1, S_2 of the PPI synchro will be disconnected from the antenna synchro and connected in parallel to the a-c line. The synchro is set up so that this locks it in position until the antenna reaches north and opens the relay coil circuit.

Very precise tie-in of the indicator deflection yoke can be obtained by means of a 36-speed synchro system and a *servoamplifier* to give torque amplification. Servosystems are discussed in Chapter 9. Here again it is necessary to use additional means to avoid ambiguity in the indicator position. An additional 1-speed synchro system accomplishes this, if provision is made for automatically switching the servoamplifier to the 1-speed system when the 1-speed error signal exceeds a certain amount (see page 283).

(*b*) The range sweep in a magnetic CRT can be rotated by applying two separate sweep currents, modulated respectively as the sine and the cosine of the desired rotation angle, to deflection coils at right angles to each other. Similarly, a set of three deflec-

tion coils can be energized by sweep currents whose maximum amplitudes vary sinusoidally with phase differences of 120°. This method requires somewhat more complicated circuits than the mechanical rotation method, but the PPI itself can be made considerably lighter and less bulky.

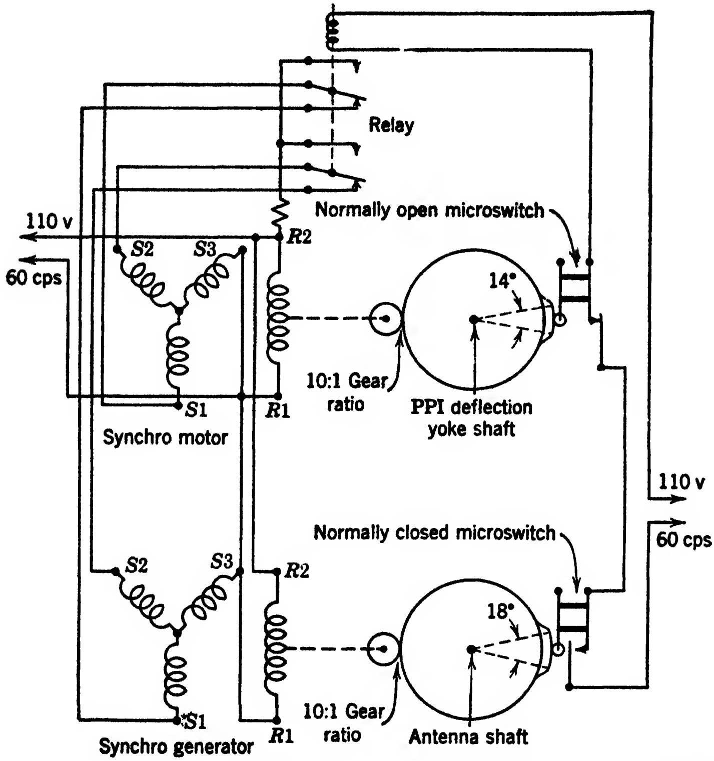


FIG. 6·9 Diagram of a 10-speed synchro system, showing switching arrangement to obtain a one-to-one correlation between antenna and deflection yoke positions.

The modulation of the sweep currents can be produced by means of rotary transformers. In the "three-phase" case ordinary three-phase synchros (Chapter 9) can be used. Figure 6·10 gives a schematic diagram for the two-phase case. A trapezoidal voltage is applied to the rotor of the transformer and induces voltages in the two-phase secondary, the maximum amplitudes of which vary sinusoidally as the rotor is turned and differ 90° in phase. A

center-tapped resistor across each secondary winding is used to give push-pull voltages relative to ground. These voltages are fed to drivers in the plate circuits of which are the deflection coils. Each deflection coil consists of four separate coils wound in a push-pull fashion, so that zero signal plate currents do not cause a deflection of the electron beam in the CRT. Because of the transformer coupling, the d-c level at the driver grids has to be re-

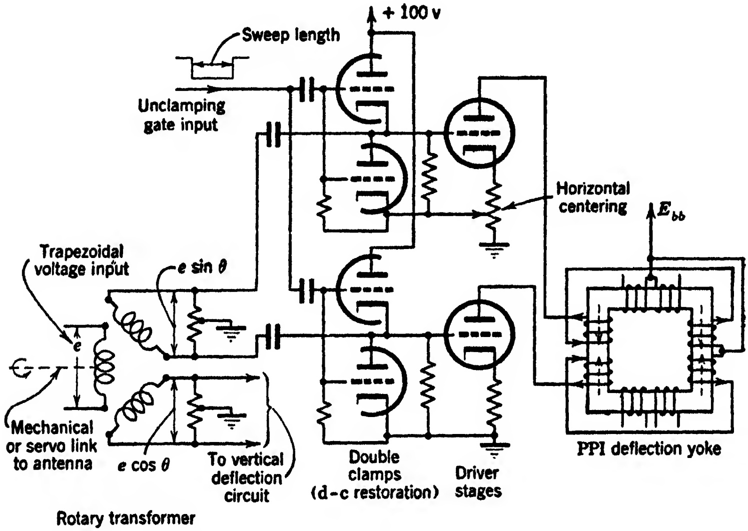


FIG. 6·10 Schematic diagram of an electromechanical PPI sweep circuit.

established; since the signals can be either positive or negative a simple diode d-c restorer (page 307) cannot be used. The arrangement shown is a double clamp (page 309) which connects the driver grids through a low impedance to ground except while the unclamping gate, equal in duration to the sweep, is applied. It is found that more reproducible clamping is obtained by connecting the plates of the upper clamp in each circuit to a low positive voltage rather than to ground.

Rotary transformers have relatively poor high frequency response, so that the rapid rise and fall of the input trapezoidal voltages are rounded out considerably. For this reason, this method is not used for fast sweeps or where the first part of the sweep has to be accurately linear.

(*c*) Another method for modulating a trapezoidal voltage is represented in Fig. 6·11. The *sine card* (cf. page 289) consists of many turns of wire on a flat card, with a contactor arm rotating about the center of the card. It is evident that if the resistance of each turn is small compared to the total resistance, the output

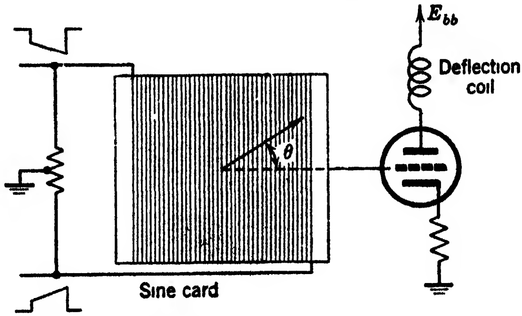


FIG. 6·11 Illustrating the use of a sine card to modulate a trapezoidal voltage.

voltage will closely approximate the desired sinusoidally modulated form.

(*d*) The range sweep in an intensity-modulated electrostatic CRT can be rotated by applying sinusoidally and cosinusoidally modulated sawtooth voltages to the two pairs of deflection plates. The modulation can be produced by a rotary transformer or a sine card.

Angle Sweeps

Angle sweeps such as are used in type B indicators are very much slower than range sweeps, and one can neglect the inductance of the deflection coils and simply regard them as resistances. Both synchros and potentiometers may be employed in circuits for developing slow angular sweeps. The use of synchros is illustrated in the B-scope azimuth sweep circuit represented in Fig. 6·12. We will consider the case in which the angle sweep is synchronized with the rotation of a radar antenna. The rotor of a synchro generator is energized by a 1500-cycle current from an oscillator, and the stator is connected to the stator of a control transformer (page 276). The amplitude of the signal induced in the rotor of the control transformer depends on the positions of

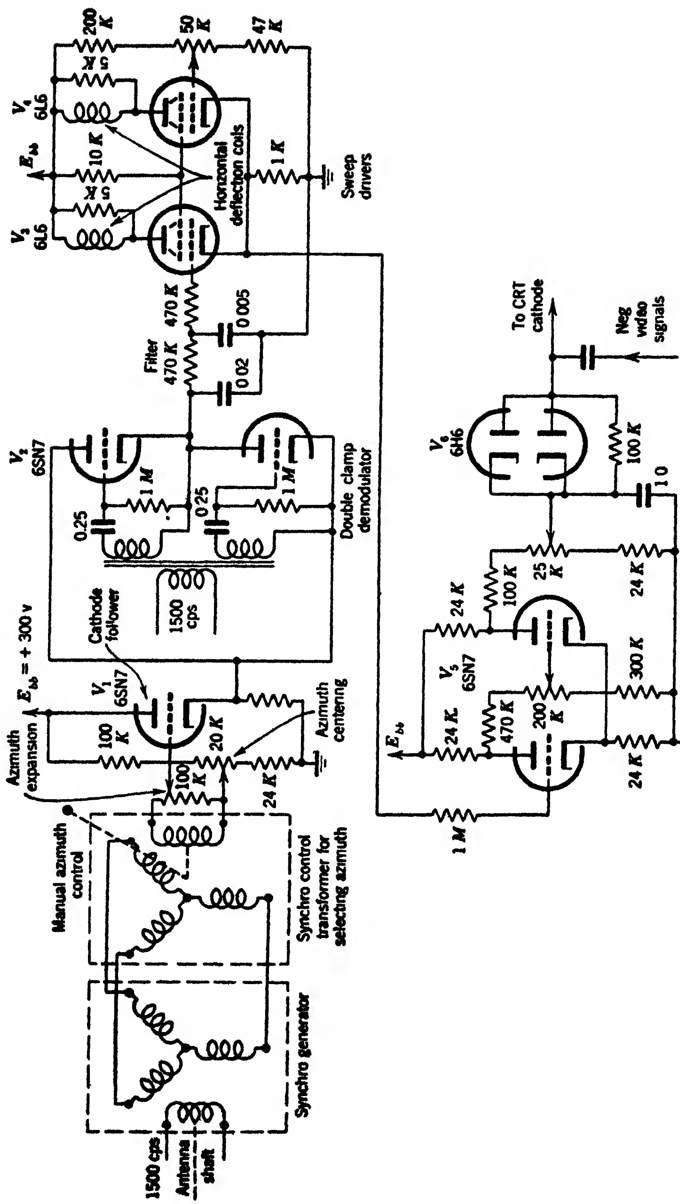


FIG. 6·12 Circuit diagram of a typical B-scope azimuth sweep circuit.

both rotors, so that the azimuth sector to be investigated can be selected by the operator. As the antenna rotates at a constant rate, the envelope of the output signal is a sine wave; this envelope is detected by the double-ended clamp V_2 (cf. page 207) and filtered, and is then applied to the grid of one of the pair of cathode-coupled drivers V_3 and V_4. The cathode coupling causes the plate current in V_4 to decrease when that in V_3 increases, and the deflection coils in the two plate circuits are arranged for push-pull operation. By means of this circuit the electron beam is deflected

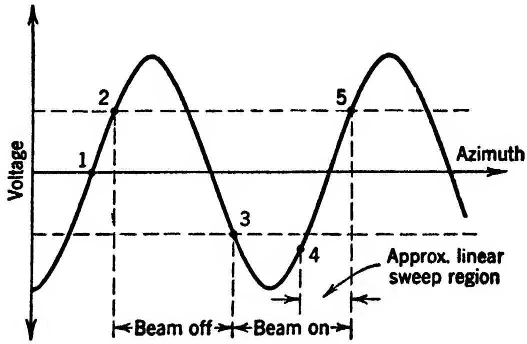


Fig. 6·13 Waveform for illustrating the azimuth sweep blanking employed in the circuit of Fig. 6·12.

back and forth across the CRT face with approximately simple harmonic motion in synchronism with the antenna rotation. The carrier frequency of 1500-cycle is used rather than 60-cycle since a smoother demodulated signal is obtained in this way.

The B-scope picture would be practically useless if the electron beam were not blanked out except during an interval of approximately 90° of antenna rotation. This blanking is produced by V_5, which is a cathode-coupled "flip-flop" circuit (see page 306). The bias on the grid of V_4 is adjusted so that when the demodulated azimuth signal (Fig. 6·13) on the cathode of V_3 and V_4 has the value at point 1 the electron beam is centered so far as the horizontal (azimuth) direction is concerned, and the bias of the grid of V_{5B} is set so that at this point V_{5B} is conducting and V_{5A} is cut off. When V_{5B} conducts, the bias on the CRT cathode permits the electron current to flow. As point 2 is approached the cutoff potential of V_{5A} is exceeded and this tube starts to conduct. This

initiates a regenerative process which quickly transfers the plate current to V_{5A}, thus applying a positive gate to the CRT cathode with resulting blanking of the beam. This situation continues until the voltage on the cathodes of V_3 and V_4 drops sufficiently to cut off V_{5A}, say at point 3. The blanking gate is then rapidly removed from the CRT, and the beam stays on until point 5 is reached. Between points 4 and 5 the sweep is reasonably linear, and if 10° electronic azimuth marks are put on the indicator the deviations from linearity are unimportant. The non-linear part of the beam-on interval, between points 3 and 4, occurs either while the beam is deflected off the tube face or is at the extreme left edge of the tube face and therefore does no harm.

As indicated in Fig. 6·12, in this particular B-scope the video signals are impressed on the CRT cathode along with the blanking gates.

Vertical Sweep in the RHI

As mentioned on page 210, the vertical sweep in a range-height indicator has an amplitude proportional to the amplitude of the horizontal sweep times the sine of the antenna elevation angle.

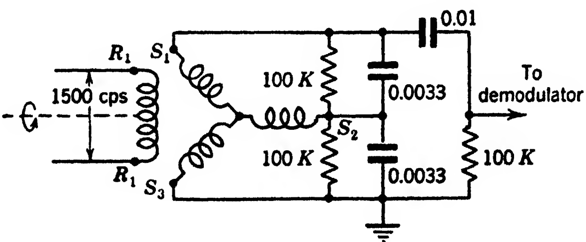


FIG. 6·14 Illustrating the use of a synchro generator to obtain a sine-modulated single-sided a-c signal.

Methods for obtaining such a modulated sweep are described above. Another method which is suitable for fast or slow sweeps is briefly described in the following paragraphs.

A voltage proportional to the sine of the elevation angle is obtained by the detection of the envelope of a 1500-cycle signal which is modulated as the sine of the elevation angle by means of a synchro generator. The circuit is essentially as shown in Fig. 6·12 through V_2, except that the synchro control transformer is

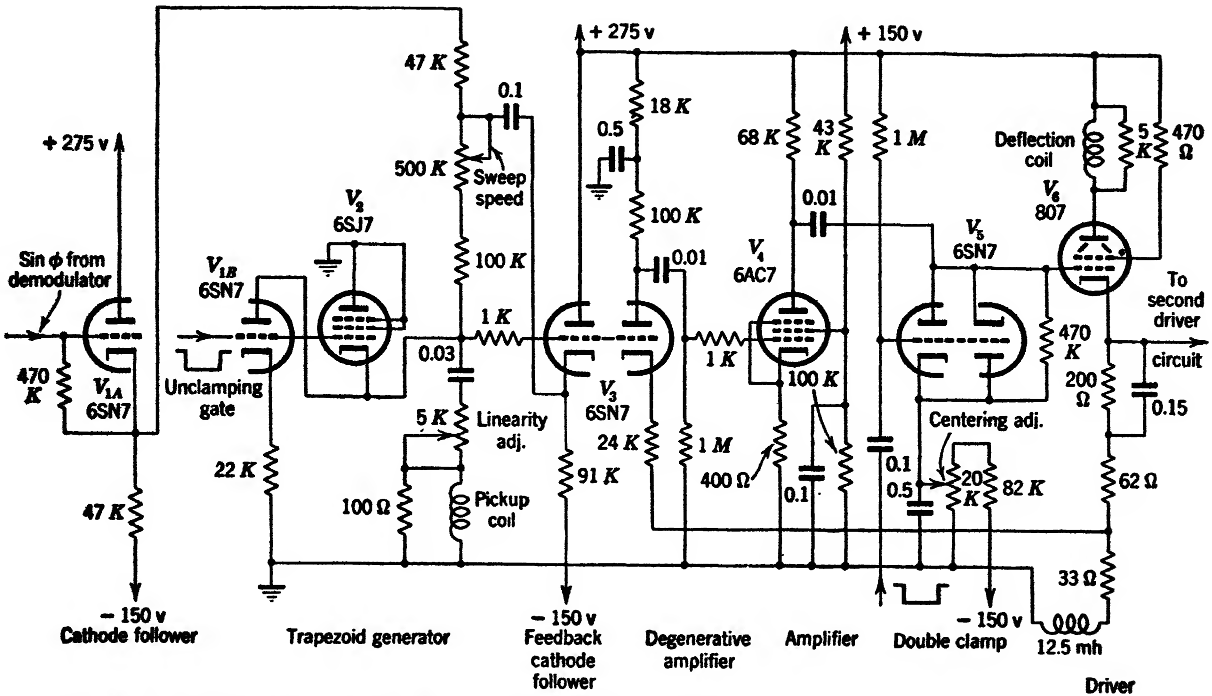


FIG. 6·15 Diagram of the vertical sweep circuit used in the range-height indicator of the "Little Abner" radar.

omitted, and the single-ended modulated signal is obtained from the stator of the synchro generator by the network shown in Fig. 6·14.

The sine voltage serves as the supply voltage for the trapezoid generator of Fig. 6·15. Since the sine voltage may go to negative values the trapezoid generator has a double-ended clamp V_{1B} and V_2. The linearity of the output is improved by the regenerative feedback supplied by the cathode follower V_{3A}, which tends to hold constant during the charging process the voltage toward which the 0.03-microfarad condenser charges (cf. the linear sweep circuit described on page 191). A small pickup coil wound on the deflection yoke is included to supply the short voltage peak necessary at the start of the trapezoid to charge the stray capacity in the deflection coil and its leads. The trapezoid voltage is amplified by a degenerative two-stage amplifier and is then fed to the sweep driver, with appropriate d-c restoration by means of a double-ended clamp. Another driver (not shown) is fed in phase opposition to give a push-pull sweep.

6·4 INDICATOR BLOCK DIAGRAMS

By way of summarizing this chapter we will discuss briefly two complete indicators, basing our discussion on the block diagrams for these indicators.

Test Scope (Types A and J Combined)

Figure 6·16 gives the block diagram for a test scope having both A- and J-presentations and using a 3-inch electrostatic tube. This scope, which carries the service designation TS-100/AP, is useful for general test purposes in connection with triggered circuits, as well as for radar system tuning and (using the circular J-sweep) range circuit calibration.

A crystal-controlled oscillator gives a sinusoidal output of 80.86 kilocycles. By means of a phase-shifting network this is changed to two sine waves 90° apart in phase, which are impressed on the deflection plates of the CRT to produce a circular sweep one revolution of which corresponds to one nautical mile.

A variable-frequency blocking oscillator produces a short gate which is applied to the cathode of a coincidence tube. Peaked

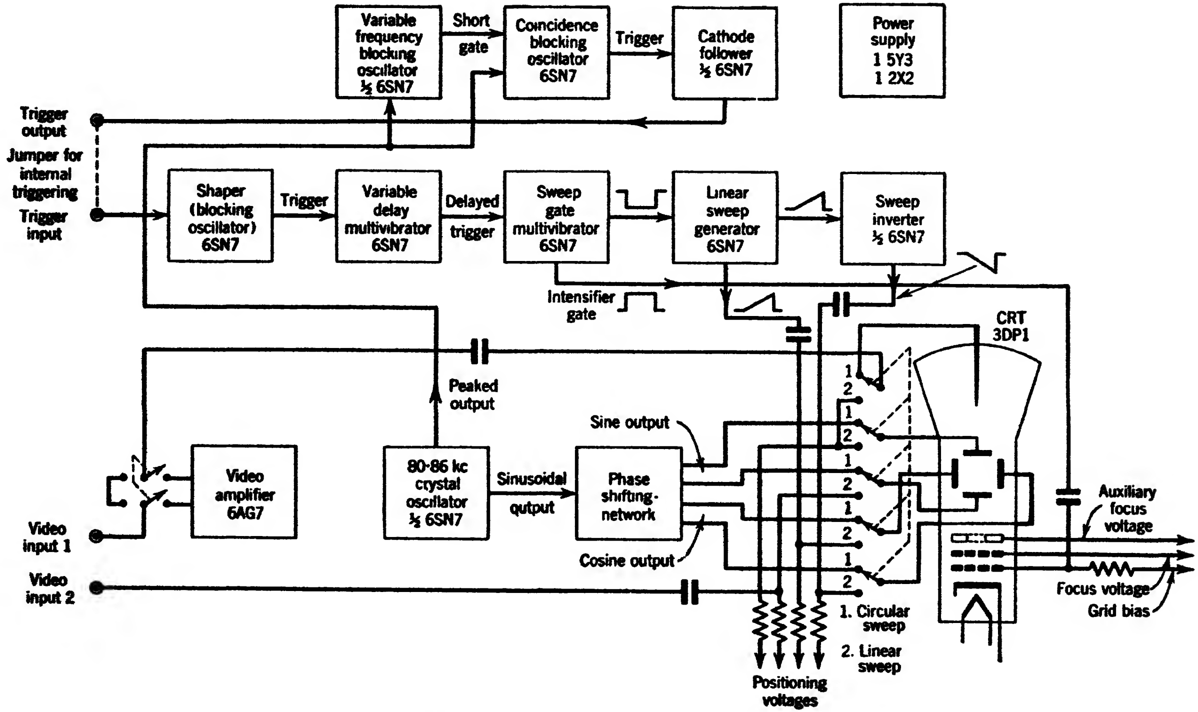


FIG. 6·16 Simplified block diagram of a 3-in. test A- and J-scope (AN/TS-100/AP).

80.86-kilocycle oscillations are applied both to the blocking oscillator and to the grid of the coincidence tube. As described on page 189, this arrangement produces a variable-frequency trigger which is accurately synchronized with the crystal oscillator output. This trigger may be used to control an external circuit; for example, a range mark generator can be triggered, and its output examined on the 1-mile circular sweep for calibration purposes.

A switch is provided for changing to linear sweep operation. Either an external trigger or the internally generated trigger can be employed. The trigger input is first used to fire a blocking oscillator. The output pulse from the blocking oscillator can be "delayed" by a conventional variable delay multivibrator (see page 182), and is then used to set off a linear sweep generator composed of a gate multivibrator, a sweep circuit employing regenerative feedback (cf. page 191), and an inverter stage to give a push-pull sweep. Various sweep speeds from about 0.2 inch per microsecond to 1 inch per millisecond can be chosen by means of a selector switch. A positive gate from the sweep gate generator is applied to the CRT grid to intensify the trace during the sweep.

A wideband video amplifier with a fixed gain of about 10 is provided. This amplifier is particularly important with the J-sweep since the deflection sensitivity with a central electrode is relatively low. The amplifier is also useful with the linear A-sweep since it allows standard 2-volt radar video signals to give a large enough deflection to be useful in tuning a radar system.

Type B Indicator

Figure 6·17 shows the block diagram for a typical B-scope. The vertical or range sweep circuit consists of a blocking oscillator trigger amplifier, a delay multivibrator, a sweep gate generator, a trapezoidal voltage generator (page 212), and two push-pull power tube drivers for supplying the relatively large deflection currents needed. D-c restoration is provided. Various sweep lengths are usually available. An appropriate intensifying signal from the sweep gate generator is applied to the first anode of the CRT. This range sweep circuit is essentially the same as that used in PPI's, RHI's, and other indicators employing magnetic tubes.

A negative gate from a multivibrator is fed to a range mark generator of the type described on page 187. This gives range

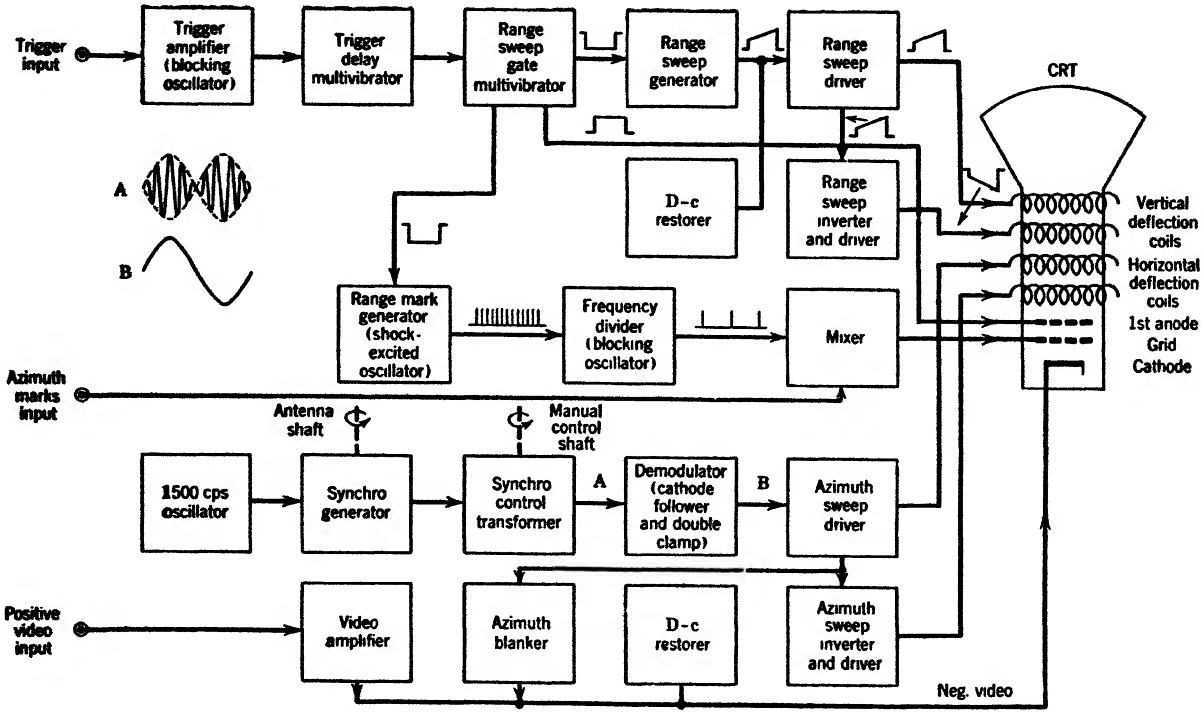


FIG. 6·17 Simplified block diagram of a type B indicator.

marks synchronized with the radar trigger; a blocking oscillator frequency divider (page 313) counts the marks down to a spacing suitable to the range being used, and the counted-down marks are fed to a mixer which combines them with azimuth marks supplied by a circuit not included in the diagram. The marks are applied to the grid of the CRT to brighten the spot at the appropriate intervals. (In some radars, both range marks and azimuth marks are produced in a central "information generator" and are cabled to the individual indicators.)

The horizontal or azimuth sweep circuit is described on page 216. A single video stage amplifies and inverts the positive video input, and the resulting negative video together with a positive azimuth blanking signal is applied to the cathode of the CRT. The maximum video gain available is sufficient to cause blooming of the CRT spot with a 2-volt input, and is controllable by the operator.

PROBLEM

6·1 Give a detailed block diagram for a range scope having a sweep speed of ⅙ in. per μsec and a "range notch" the position of which relative to the radar system trigger is determined by means of a precision delay circuit. The notch is a deflection in the CRT trace a few microseconds long, and range to a target is determined by adjusting the position of the notch so that the target echo takes a standard position within it. Indicate all significant waveforms, showing their proper time relations. Suggest types of circuits which could be used for each of the important operations in this scope.

CHAPTER 7

TUNED AMPLIFIERS

The majority of receivers used in microwave radars are of the *superheterodyne*[1] type, in which the received r-f signals are converted to an intermediate frequency by being mixed with the output of a local oscillator in a non-linear circuit element. The resulting intermediate frequency (i-f) signals are then amplified by a high gain *tuned* amplifier before being finally converted to video signals for application to the radar indicators. The wartime development of i-f amplifiers for this application involved some advances, particularly with respect to achieving wide bandwidths for the amplification of pulsed signals, which it is the purpose of the present chapter to discuss briefly. We will include, also, a brief discussion of one type of amplifier tuned for audio frequency signals.

7·1 SINGLE-TUNED AMPLIFIER STAGES; BANDWIDTH AND GAIN

A single-tuned amplifier may be represented by the schematic diagram of Fig. 7·1. In a typical i-f stage, the resistance R (Fig. 7·2) is made up of the load resistor R_L, the equivalent shunt resistances of L and C, the plate resistance of the first tube, and the input resistance of the second tube, all in parallel; the capacity C is composed of the shunt capacity of L, and tube, wiring, and socket capacities. In most cases R is practically equal to R_L. The inductance is chosen to resonate with the capacitance at the tuned frequency $f_0 = \omega_0/2\pi$.

[1] A discussion of microwave receivers is given in Chapter 8.

The impedance of the coupling network of Fig. 7·2 is

$$Z(\omega) = \frac{R}{1 + jQ\left(\dfrac{\omega}{\omega_0} - \dfrac{\omega_0}{\omega}\right)} \tag{7·1}$$

where $\omega = 2\pi f$, $\omega_0 = 1/\sqrt{LC}$, and $Q = \omega_0 RC$. If V_1 is a pentode

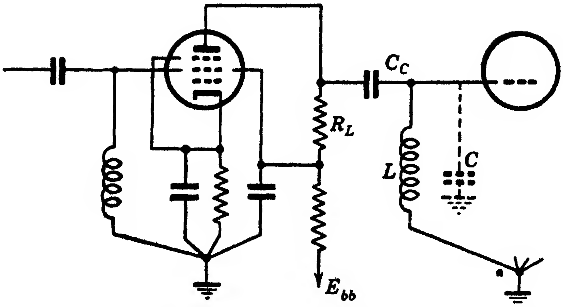


FIG. 7·1 A typical single-tuned intermediate frequency amplifier stage.

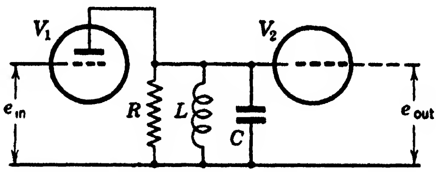


FIG. 7·2 Schematic representation of a single-tuned stage neglecting d-c returns.

of transconductance g_m, the voltage gain from the grid of V_1 to that of V_2 is

$$A(\omega) = g_m Z(\omega) \tag{7·2}$$

since for a pentode $r_p \gg |Z(\omega)|$. At band-center ($\omega = \omega_0$),

$$A(\omega_0) = g_m R \tag{7·3}$$

The bandwidth of an amplifier is usually defined as the separation of the frequencies at which the voltage gain is $1/\sqrt{2}$ of its value at band center; these frequencies are the so-called half-power or 3-decibel points. This bandwidth, which is very convenient mathematically, is nearly the same as the noise bandwidth

defined on page 246 in those cases, such as several cascaded amplifier stages, which are of interest to us here. Solution of the equation

$$\frac{|A(\omega)|}{A(\omega_0)} = \frac{1}{\sqrt{2}}$$

gives for the 3-decibel angular frequencies

$$\omega_1 = \frac{1}{2Q}\left(-1 + \sqrt{1 + 4Q^2}\right)$$

$$\omega_2 = \frac{1}{2Q}\left(1 + \sqrt{1 + 4Q^2}\right)$$

so that the bandwidth is

$$\Delta f = f_2 - f_1 = \frac{f_0}{Q} = \frac{1}{2\pi RC} \tag{7·4}$$

The product of the band center gain and the bandwidth, the *gain-bandwidth product* (cf. page 150),

$$A(\omega_0)\Delta f = \frac{g_m}{2\pi C} \tag{7·5}$$

is a quantity of importance. It depends on the tube type, except for that part of C which is due to wiring. With good wiring C can be assigned a fairly constant value for each tube type, so that the gain-bandwidth product can be used as a figure-of-merit for each tube type. It has the value of approximately 57 for 6AC7's ($g_m = 0.009$ mho, $C = 25$ micromicrofarads), and 66 for 6AK5's ($g_m = 0.005$ mho, $C = 12$ micromicrofarads), if f is expressed in megacycles. Thus if a single-tuned 6AC7 stage has a bandwidth of 5 megacycles, its voltage gain is limited to approximately 11.

Following the treatment given by Wallman,[2] it is convenient to make the following substitutions in the above equations:

$$\beta = \frac{2\pi C}{g_m} A(\omega) \tag{7·6}$$

$$d = \text{dissipation factor} = \frac{1}{Q} \tag{7·7}$$

[2] H. Wallman, Radiation Laboratory Report 524, Feb. 1944.

If in addition the ratios f/f_0 and f_0/f are replaced by f and $1/f$ respectively, which amounts to taking the center frequency equal to unity, equation 7·2 takes the simple form

$$\beta = \frac{1}{d + j\left(f - \frac{1}{f}\right)} \tag{7·8}$$

which may be termed the *normalized gain function* for a single-tuned circuit. The *amplitude function* is

$$|\beta| = \frac{1}{\sqrt{d^2 + \left(f - \frac{1}{f}\right)^2}} \tag{7·9}$$

and the *phase angle* is

$$\phi = \tan^{-1} \frac{1}{d}\left(f - \frac{1}{f}\right) \tag{7·10}$$

At the band center

$$|\beta|_{max} = \frac{1}{d} \tag{7·11}$$

The *gain-bandwidth factor*, $|\beta|_{max}\,\Delta f$, is the gain-bandwidth product (equation 7·5) expressed in units of $g_m/2\pi C$, and obviously has the value unity.

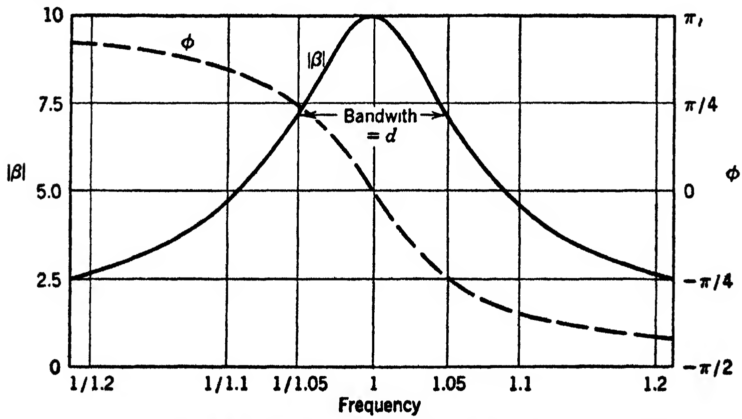


FIG. 7·3 Amplitude and phase angle of the gain function of a single-tuned stage having dissipation factor = 0.1.

Figure 7·3 shows the quantities $|\beta|$ and ϕ as functions of f for a single-tuned stage having $d = 0.1$, a value typical for i-f ampli-

fiers using 6AC7's or 6AK5's and tuned to approximately 30 megacycles.

Cascaded Single-Tuned Stages

If n identical synchronously[3] tuned single-tuned stages are cascaded, the overall gain function is simply the *product* of the individual gain functions, since the tubes effectively isolate the coupling networks from each other. Thus the gain function for such a system is

$$\beta_n \equiv \beta^n = \frac{1}{\left[d + j\left(f - \frac{1}{f}\right)\right]^n} \tag{7·12}$$

The overall bandwidth is

$$\Delta f_n = (2^{1/n} - 1)^{1/2}\,\Delta f \tag{7·13}$$

Values of $\Delta f_n/\Delta f$ are given in Table 7·1. It is evident that the overall bandwidth becomes smaller relative to the single-stage bandwidth as the number of stages increases. Since the product of stage gain and stage bandwidth is constant, this means that, for a given tube type and a given overall gain, there is a maximum bandwidth which can be obtained no matter how many stages are employed. For 6AC7's and an overall gain of 10^5, this maximum occurs for 23 stages and is only 6 megacycles. A larger number of tubes giving the same overall gain would lead to a smaller overall bandwidth.

TABLE 7·1 RATIO OF OVERALL BANDWIDTH OF CASCADED SINGLE-TUNED STAGES TO SINGLE-STAGE BANDWIDTH

n	$\Delta f_n/\Delta f$	n	$\Delta f_n/\Delta f$
1	1	6	0.35
2	0.64	7	0.32
3	0.51	8	0.30
4	0.44	9	0.28
5	0.39	∞	0

The appropriate figure of merit for a multistage amplifier, analogous to the gain-bandwidth factor for a single stage, is the

[3] Synchronously tuned stages are stages which are tuned to the same frequency.

stage gain-overall bandwidth factor, which in the case of n synchronous single-tuned stages is $(2^{1/n} - 1)^{1/2}$, the quantities listed in Table 7·1 [since the stage gain $= (1/d^n)^{1/n}$].

The response of a single-tuned stage and of six cascaded synchronously tuned stages of different bandwidths to a pulse of i-f oscillations having a rectangular envelope is illustrated in Fig. 5·3.

7·2 STAGGER TUNING OF SINGLE-TUNED STAGES

The data of Table 7·1 indicate that it is not practical to achieve bandwidths in excess of 2 or 3 megacycles in high gain i-f amplifiers with synchronously tuned single-tuned stages. One of the most satisfactory methods of increasing the bandwidth without decreasing the overall gain is to "stagger" the tuning of the individual stages (that is, to cascade stages which are tuned to different frequencies). It can be shown [4] that the problem reduces to adjusting the center frequency and the dissipation factor of each stage so that the amplitude function has the form

$$|\beta_n| = \frac{1}{\left[\delta^{2n} + \left(f - \frac{1}{f}\right)^{2n}\right]^{1/2}} \tag{7·14}$$

instead of the form

$$|\beta_n| = \frac{1}{\left[d^2 + \left(f - \frac{1}{f}\right)^{2}\right]^{n/2}} \tag{7·15}$$

corresponding to equation 7·12. An amplifier with the amplitude function 7·14 is called an *exact-staggered n-tuple;* the maximum gain is

$$|\beta_n|_{\max} = \frac{1}{\delta^n} \tag{7·16}$$

and the stage gain [5] is $1/\delta$, so that the stage gain-overall bandwidth factor is unity, as it is for a single single-tuned stage.

[4] H. Wallman, Radiation Laboratory Report 524, Feb. 1944; *Electronics,* **21,** 100 (May 1948).

[5] In a staggered n-tuple amplifier the gains of all the stages are not equal. The stage gain is defined as the geometric mean of the individual gains at overall band center.

The synthesis of the amplitude function 7·14 is an important result since it shows how to cascade simple single-tuned stages to get the requisite high gain without any sacrifice in overall bandwidth. This synthesis leads to the results given in Table 7·2 for $n = 2$ and 3.

TABLE 7·2 SYNTHESIS OF EXACT-STAGGERED n-TUPLES

Center frequency $= f_0$; overall bandwidth $= \Delta f$; $\Delta f/f_0 = \delta$

n	Component Single-Tuned Stages
2 (staggered pair)	Two stages of dissipation factor d, tuned to af_0 and f_0/a where $$d^2 = \frac{4 + \delta^2 - \sqrt{16 + \delta^4}}{2}$$ and $$\left(a - \frac{1}{a}\right)^2 + d^2 = \delta^2$$
3 (staggered triple)	Two stages of dissipation factor d, tuned to f_0/a. One stage of dissipation factor δ, tuned to f_0 where $$d^2 = \frac{4 + \delta^2 - \sqrt{16 + 4\delta^2 + \delta^4}}{2}$$ and $$\left(a - \frac{1}{a}\right)^2 + d^2 = \delta^2$$

It will be noted that the individual stage gains are not specified in the table. The maximum gain which can be obtained from each stage is limited by the gain-bandwidth product of the tube type used, so that the maximum overall gain of the amplifier is of course also limited. However, it is an important fact that the *shape* of the overall amplitude curve is independent of the gain of the individual stages, so that any of the stages may be gain-controlled (page 260) at will.

To obtain high gain amplifiers having bandwidths not much greater than 10 megacycles, with $f_0 = 30$ megacycles, one usually cascades three or four sets of identical staggered pairs or triples. If the individual stages are adjusted, according to the theoretical requirements, to the proper dissipation factors [6] and center fre-

[6] It is found that the load resistors have to be somewhat larger than calculated because of the loading effect due to tube input resistance and coil resistance. The values are chosen experimentally to give the proper dissipation factors.

quencies, the amplifier is bound to have the desired properties, provided there is no interaction between the stages. The overall amplitude function for m exact-staggered n-tuples is

$$|\beta_{mn}| \equiv |\beta_n|^m = \frac{1}{\left[\delta^{2n} + \left(f - \frac{1}{f}\right)^{2n}\right]^{m/2}} \tag{7·17}$$

Therefore the bandwidth is

$$\Delta f_{mn} = (2^{1/m} - 1)^{1/2n}\,\delta \tag{7·18}$$

From this expression one can see what value of δ is necessary to give the desired overall bandwidth.

Figure 7·4 illustrates how the individual stages in a staggered triple combine to form a wide, flat-topped amplitude curve. The properties of the individual stages are indicated in the figure. For convenience, the stage gain at band center has been set equal to unity; in this case, the result of stagger tuning which should be noted is that the overall bandwidth is 9 megacycles, whereas if three cascaded stages like the center one were synchronously tuned the overall bandwidth would be only 4.6 megacycles. Obviously, if the stage gain were greater than unity the overall curve would no longer coincide with the curve for the center stage at the band center.

Figure 7·5 shows the amplitude curve for an amplifier composed of three exact staggered triples centered at 30 megacycles, and having a bandwidth of 10.6 megacycles. This amplifier, designed by H. Logemann of the Radiation Laboratory, has a gain of 108 decibels, or a stage gain of 12 decibels. The figure illustrates the excellent agreement between predicted and observed behavior to be expected in setting up a stagger-tuned amplifier.

The sharp "skirts" evident in Fig. 7·4 and 7·5 are characteristic of stagger-tuned amplifiers. The increased 3-decibel bandwidth is obtained at the expense of gain outside the band, so that the drop in gain at the edges of the band is rapid. This rapid drop is inevitably accompanied by rapid changes in phase shift, so that a signal which includes a wide range of frequencies suffers considerable distortion on passing through the amplifier. Wallman [7] has shown that for high-Q staggered n-tuples, for which the normalized

[7] H. Wallman, Radiation Laboratory Report 524, Feb. 1944.

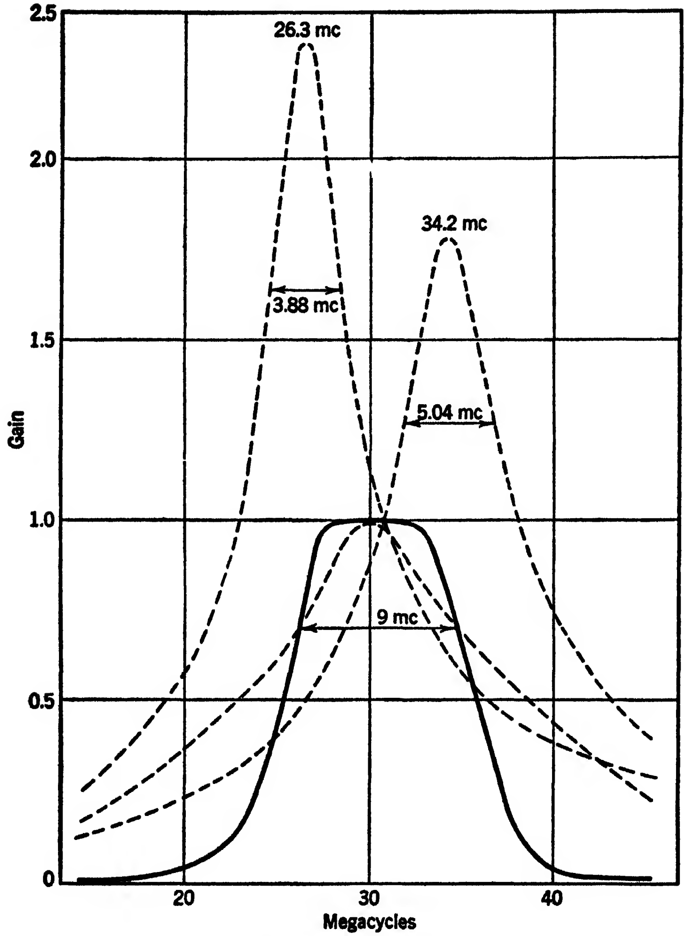


FIG. 7·4 The flat-topped bandpass characteristic (solid curve) is the product of the three single-tuned characteristics (dotted curves) of an exact-staggered triple.

amplitude function is of the type $1/\sqrt{1 + x^{2n}}$, where $x = f - f_0$, the response to a step voltage is such that overshoots of 0, 4.3, 8.1, 10.9, 12.8, and 14.3 per cent are obtained with $n = 1, 2, 3, 4, 5$, and

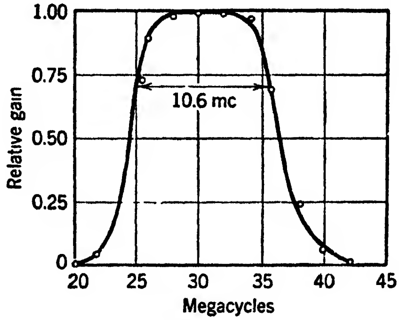


FIG. 7·5 The amplitude curve for a nine-stage amplifier consisting of three exact-staggered triples and having an overall gain of 108 db. The curve is the predicted curve, and the circles represent measured values.

6, respectively. It is to be noted that this same limitation arises with other types of amplifiers having stage gain-overall bandwidth factors comparable to that of the staggered n-tuple.

7·3 DOUBLY TUNED AMPLIFIER STAGES

More complicated coupling networks than that of Fig. 7·1 find application in tuned amplifiers. A doubly tuned transformer-

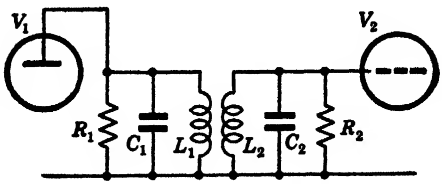


FIG. 7·6 Schematic diagram, neglecting d-c returns, of a doubly tuned amplifier stage.

coupled network is shown schematically in Fig. 7·6. In most cases L_1 and L_2 are chosen to resonate with C_1 and C_2, respectively, at the same frequency. In amplifiers tuned to high frequencies

C_1 and C_2 are usually the distributed capacities of the primary and secondary circuits. It can be shown that for high-Q circuits having equal primary and secondary Q's adjusted to critical coupling,[8] the gain-bandwidth factor is $\sqrt{2}$; if one side of the circuit is loaded, the factor is increased to 2. Because of this fact, doubly tuned coupling networks are widely used. In general, an amplifier composed of cascaded doubly tuned stages is considerably more difficult to adjust than is one of equal stage gain-overall bandwidth factor made up of staggered pairs or triples.

7·4 TUNED AUDIO AMPLIFIERS EMPLOYING THE TWIN-TEE NETWORK

The principles of stagger tuning outlined above apply as well to audio amplifiers as they do to i-f amplifiers, provided only that the gain function for each of the staggered units approximates that for a single-tuned circuit. The use of L-C tuned circuits, especially at low audiofrequencies, is frequently inconvenient because of the difficulty of obtaining suitable components and because of the danger of magnetic pickup. It has been found [9] that a gain function sufficiently close in character to that of a single-tuned L-C circuit can be obtained by using frequency-dependent inverse feedback supplied by an R-C network such as a Wien bridge or a twin-tee filter.[10] This fact has been applied [9] to the design of band-pass audio amplifiers having amplitude functions difficult to obtain by other means.

The twin-tee network is shown in Fig. 7·7, together with its amplitude and phase characteristics, derived from the expression for the transmission characteristic:

[8] H. Wallman, Radiation Laboratory Report 524, Feb. 1944, p. 7. See also J. G. Brainerd, G. Koehler, H. J. Reich, and L. F. Woodruff, *Ultra-High Frequency Techniques*, D. Van Nostrand, 1942, p. 143; F. E. Terman, *Radio Engineers' Handbook*, McGraw-Hill Book Co., 1943, pp. 154 and 436.

[9] (*a*) R. Walker and H. Fleisher, Radiation Laboratory Report 737, May 1945; (*b*) R. M. Ashby, F. W. Martin, and J. L. Lawson, Radiation Laboratory Report 914, March 1946.

[10] H. H. Scott, *Proc. I.R.E.*, **26**, 226 (1938).

$$\beta = \frac{E_{\text{out}}}{E_{\text{in}}} = \frac{1}{1 - 4j\dfrac{1}{\dfrac{\omega}{\omega_0} - \dfrac{\omega_0}{\omega}}} \tag{7·19}$$

where $\omega_0{}^2R^2C^2 = 1$. Since the output of a perfect twin-tee network vanishes at ω_0, an amplifier with a negative feedback loop containing such a filter will have a gain which is a maximum at ω_0.

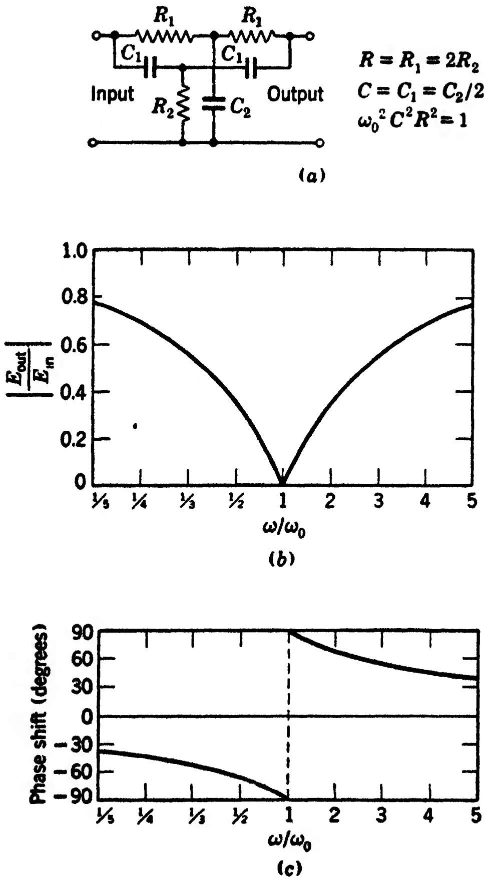


FIG. 7·7 The balanced twin-tee network and its amplitude and phase characteristics. (a) Diagram of the network; (b) amplitude transmission characteristic; (c) phase transmission characteristic.

Consider the simple case in which the output E_2 of an amplifier having a *real* voltage gain μ (Fig. 7·8) is fed to a twin-tee network,

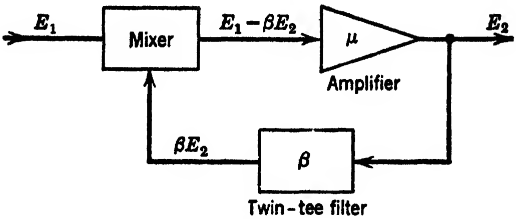


FIG. 7·8 Block diagram of an amplifier having a twin-tee filter in a feedback loop. If the feedback is degenerative, the gain of the amplifier is maximum at the null frequency of the twin-tee.

and the output of the latter is subtracted by a suitable mixer from the input signal E_1. We have

$$E_2 = \mu(E_1 - \beta E_2)$$

so that the gain A of the system is

$$A = \frac{E_2}{E_1} = \frac{\mu}{1 + \beta\mu} \tag{7·20}$$

At $\omega = \omega_0$, $\beta = 0$ and $A_0 = \mu$. If we define the Q of the circuit in the same way as for a single-tuned circuit, that is

$$Q = \frac{\omega_0}{\omega_2 - \omega_1} \tag{7·21}$$

where ω_1, ω_2 are the angular frequencies corresponding to half-power output, we can evaluate Q by finding the frequencies for which

$$\frac{|A|^2}{A_0^2} = \frac{1}{2}$$

In this way it is found that, if $\mu \gg 1$,

$$Q = \frac{\mu}{4} \tag{7·22}$$

The characteristic of the twin-tee itself is such that the half-power bandwidth is $4\omega_0$. Thus we see that the "Q" of the twin-tee is

increased by the factor μ when it is used in a feedback loop. Since values of μ of several hundred are easily obtained, it is evident that high values of Q may be realized.

It is interesting to point out that by a suitable small off-balancing of the twin-tee network $[R_2 < (R_1/2)]$, so that a null is not obtained and the output near ω_0 is regenerative, a moderately stable circuit can be obtained which has a Q of many thousands even at very low audiofrequencies.

It is readily seen that if the output of the system in Fig. 7·8 is taken from the *output* side of the twin-tee, i.e., is equal to βE_2, the gain varies between unity (if $\mu \gg 1$) at frequencies far from $\omega_0/2\pi$ and zero at $f_0 = \omega_0/2\pi$. Thus the circuit can be used as a *rejective* amplifier, and we again have a half-power bandwidth given by

$$\omega_2 - \omega_1 = \frac{4\omega_0}{\mu} \tag{7·23}$$

The circuits[11, 12] shown in Fig. 7·9 illustrate two practical methods of incorporating a twin-tee network in a feedback loop. The first of these circuits is composed of a so-called "cascode" amplifier with plate-to-cathode direct coupling. The latter circuit has the important feature of using direct coupling throughout the feedback loop so that there is strong degeneration at zero frequency, thereby giving excellent circuit stability. Stagger-tuned amplifiers made up of several such sections have been found to have overall amplitude functions approaching very closely to what one would expect on the basis of the treatment given earlier in this chapter. Truly remarkable gain characteristics can be obtained; for example, a characteristic centered at 16.5 cycles, essentially flat between 15 and 18 cycles, and 45 decibels down at 13 and 21 cycles was obtained[11] by the use of a staggered quintuple.

For some purposes it may be desired to have a selective amplifier of this general type which at the same time is linear and very stable (that is, has a performance only very slightly affected by supply voltages and tube characteristics). Such a result[12, 13] is ob-

[11] R. Walker and H. Fleisher, Radiation Laboratory Report 737, May 1945.

[12] R. M. Ashby, F. W. Martin, and J. L. Lawson, Radiation Laboratory Report 914, March 1946.

[13] J. M. Sturtevant, *Rev. Sci. Instruments*, 18, 124 (1947).

tained by paralleling the frequency-dependent feedback with frequency-independent feedback, so that considerable degeneration will be present even at ω_0. In this way a circuit is obtained whose band center gain and Q are determined almost entirely by the values of a few resistors.

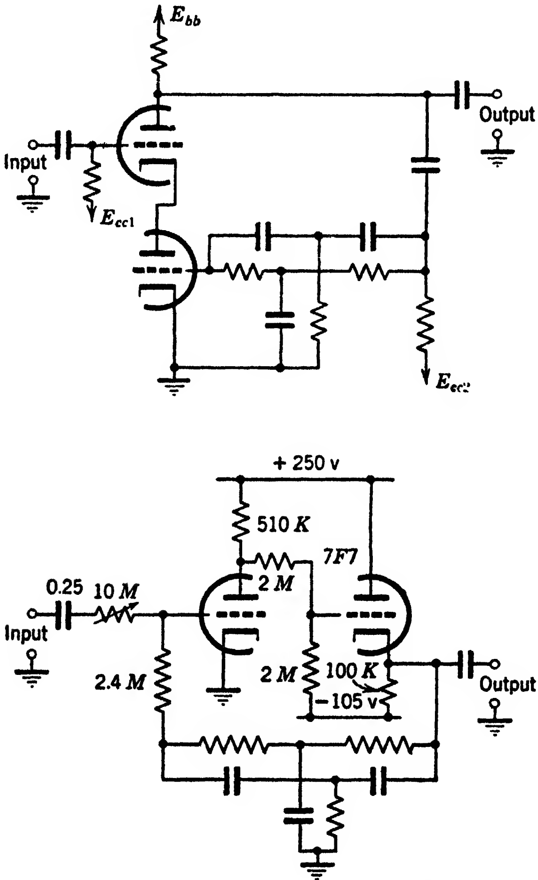


FIG. 7·9 Circuit diagrams illustrating two methods which have been used for including a twin-tee network in a feedback amplifier.

PROBLEMS

7·1 Derive equation 7·1 expressing the impedance of the coupling network of a simple single-tuned amplifier stage.

7·2 Derive equations 7·13 and 7·18 from the respective functions given in equations 7·12 and 7·17.

7·3 Show that the π-section of Fig. 7·10 is equivalent to the T-section if the impedances are related by the expressions

$$\frac{Z_1}{Z'Z'''} = \frac{Z_2}{Z'Z''} = \frac{Z_3}{Z''Z'''} = \frac{1}{Z'} + \frac{1}{Z''} + \frac{1}{Z'''}$$

(Assume an impedance Z across the output terminals of each section and compute the resulting impedances looking into each section.)

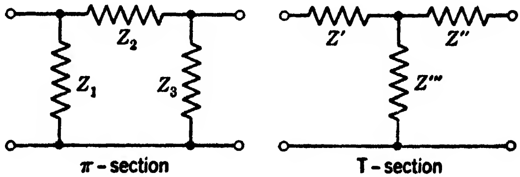


FIG. 7·10 The π- and T-sections are equivalent when

$$\frac{Z_1}{Z'Z'''} = \frac{Z_2}{Z'Z''} = \frac{Z_3}{Z''Z'''} = \frac{1}{Z'} + \frac{1}{Z''} + \frac{1}{Z'''}$$

7·4 Use the result of Problem 7·3 to derive the expression for the transmission of a twin-tee network given in equation 7·19.

7·5 Compute the tuned frequencies and bandwidths of the three stages of an exact-staggered triple with a center frequency of 100 cps and a bandwidth of 30 cps. Sketch the three individual and the overall amplitude functions.

CHAPTER 8

AMPLIFICATION OF VERY WEAK SIGNALS

The amplification of very weak a-c signals involves complications which are not encountered with signals having amplitudes of the order of millivolts or volts. It might be supposed that one could handle a signal of arbitrarily small amplitude simply by cascading a sufficient number of amplifier stages of the types discussed in Chapters 5 and 7, or in the very voluminous literature on amplifiers. However, a fundamental limitation is imposed here because electrical currents are carried by electrons which, at ordinary temperatures, always have random thermal motions superimposed on any steady motion resulting from the application of an electric field. These thermal motions give rise to finite random voltages called *noise*, and in any amplifier the useful limit of gain is reached when the output contains such noise voltages in appreciable amplitude.

A secondary complication in the design of amplifiers for handling very weak signals arises from the ever-present danger of accidental coupling between high level and low level stages. Such coupling, which may result from the common power supply, or from ground currents in the chassis, or from a variety of other causes, will in general lead at best to unpredictable performance, and at worst to oscillation.

Noise also limits the useful amplification which can be applied to d-c signals. In most applications involving small d-c signals it is found to be more satisfactory to convert the d-c signal to an a-c

signal, by means of some sort of mechanical switching device,[1] before amplification, rather than to run into the difficulties inherent in a high gain d-c amplifier.

The signal returned to a radar by scattering from a small target such as an aircraft at a distance of scores of miles represents a peak power of an extremely small fraction of a watt (cf. page 322). It is obvious that, since the noise in the receiving system defines the smallest signal which can be detected, it is of the greatest importance to investigate carefully all sources of noise with a view to minimizing their effect. The handsome returns which may result from such investigation are well illustrated by the development in microwave receivers which took place between 1942 and 1945. By constant attention to the noise problem, crystal mixers (page 251) and intermediate frequency amplifiers were improved in this interval so much that in 1945 signals with only about one-sixth the power required for detection by the earlier receivers could be detected. Obviously this improvement in the receiver increases the overall performance of a radar as much as would a sixfold increase in the transmitter power, and with an incomparably smaller increase in the cost, weight, and complexity of the system.

Our consideration in this chapter of the problems inherent in the amplification or detection of small signals will in the first place be concerned with a discussion of noise and its origins. We shall then discuss in some detail the various types of systems which have found application in the detection of weak microwave signals. In this connection, some attention will be given to various factors which affect the visibility, on a cathode ray tube screen, of the pulsed video signals derived from microwave signals, in the presence of noise and in the presence of other types of interference.

8·1 NOISE

Thermal Noise in a Conductor

The random thermal motions possessed by the electrons in a conductor at temperatures above absolute zero give rise to random voltages between the terminals of the conductor, in addition to

[1] M. D. Liston, C. E. Quinn, W. E. Sargeant, and G. G. Scott, *Rev. Sci. Instruments*, **17**, 194 (1946).

any voltage resulting from a uniform motion due to an applied electric field. Johnson [2] and Nyquist [3] have shown that these voltage fluctuations have a mean square value in the frequency range f_1 to f_2 given by

$$\overline{E^2} = 4kT \int_{f_1}^{f_2} R \, df \tag{8·1}$$

where k is Boltzmann's constant (1.374×10^{-23} joule per degree Kelvin), T is the absolute temperature (degrees Kelvin), and R is the resistive component of the impedance of the conductor. If R is independent of frequency, integration gives

$$\overline{E^2} = 4kTR(f_2 - f_1) \tag{8·2}$$

It is evident that the noise developed in a pure resistance is distributed uniformly over the entire frequency spectrum, since only the difference $f_2 - f_1$ occurs in equation 8·2.

It is convenient to consider a resistor R as a generator characterized by an *available noise power* P_n, available power being defined as the *maximum* amount of power which can be transferred from the generator to a load. Since the maximum power transfer takes place when the load impedance is the complex conjugate of the generator impedance, the available noise power is

$$P_n = kT(f_2 - f_1) \tag{8·3}$$

The available noise power has the convenient property that it is independent of resistance. At ordinary temperatures (300°K) it has the value 4×10^{-15} watt per megacycle.

Granular resistance units such as carbon resistors have available noise powers far exceeding the value indicated above when they are carrying direct currents. Such resistors should therefore never be used in the low level stages of amplifiers.

Tube Noise

Noise is developed in vacuum tubes by several causes.

(*a*) The so-called *shot effect* is due to randomness in the time and velocity of departure of electrons from the cathode. It becomes

[2] J. B. Johnson, *Phys. Rev.*, **32**, 97 (1928).

[3] H. Nyquist, *ibid.*, **32**, 110 (1928).

most pronounced under conditions of temperature-limited emission (that is, when all the electrons emitted by the cathode reach the plate). It is easy to see in a rough way why this is so, for under conditions of space charge limitation of the tube current a large excess of electrons is emitted by the cathode which do not reach the plate and which cause a smoother statistical average of tube current because of their interactions with the electrons which finally do reach the plate.

As might be expected, the shot-effect noise power is distributed evenly over the frequency spectrum, as is thermal agitation noise; this fact, together with the fact that the shot-effect noise *power* is directly proportional to the first power of the tube current, makes it possible to construct convenient *noise generators* for measurement and testing purposes from temperature-limited diodes.

(*b*) *Partition noise* results from random variations in the distribution of the total cathode current between the plate and any other electrodes, such as the screen grid of a pentode, which are positive with respect to the cathode. Because of partition noise, the total noise power output of a pentode is some three to five times as large as that of a triode with the same amplification.

(*c*) *Induced grid noise* results from currents induced in the grid circuit by shot-effect variations in the electron stream.

Noise from all these sources is essentially evenly distributed over the frequency spectrum and may therefore be expressed in terms of an equivalent *noise resistance.*

Tube noise in the low audiofrequency range also arises from the so-called *flicker effect* due to random changes in the emission characteristics of the cathode. Noise from this source may be many times larger than that from all other sources combined at very low frequencies. In our further discussion of noise we will restrict attention to those types which are uniformly distributed over the entire spectrum.

Noise Figures

The output of an ideally perfect amplifier contains no noise power over and above that due to amplification of the noise power supplied by preceding elements to its input terminals. We can therefore establish a figure of merit for an amplifier, which expresses its closeness of approach to the ideal case as follows. Suppose that

there is connected to the input terminals of the amplifier an impedance equal to the complex conjugate of the input impedance of the amplifier, and the noise power developed in a properly matched load at the output terminals is measured. Then we may take as our figure of merit the *noise figure* of the amplifier, defined as the total available noise power output divided by the available noise power output over the pass band of the amplifier due to the input impedance:

$$NF \equiv \frac{\text{available noise power output}}{kTG_{\max}\,\Delta f} \tag{8·4}$$

Here G_{max} is the maximum *power* gain, and Δf is the bandwidth defined by the expression

$$\Delta f \equiv \frac{1}{G_{\max}} \int_0^\infty G\,df \tag{8·5}$$

It was mentioned on page 227 that the bandwidth defined in this way is nearly equal to the half-power or 3-decibel bandwidth for the types of band pass characteristics usually encountered in multistage amplifiers.

It will be necessary for some of our later discussion to see how the overall noise figure of a system of cascaded units depends on the individual noise figures. Consider the case of two four-terminal units connected as shown in Fig. 8·1. The ideal signal generator

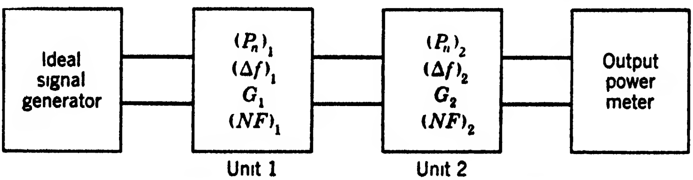


FIG. 8·1 Schematic arrangement for analyzing the noise output from two cascaded units, and for illustrating the determination of the overall noise figure of two such units.

is a device which gives an output signal of known power which can be varied at will and can be reduced in value to the theoretical lower limit equal to the available noise power of its output resistance. The output signal may be either noise or a sine wave of the frequency which is amplified most by the two four-terminal units in cascade. It is assumed that all input and output imped-

ances are properly matched. For each of the four-terminal units we have, by definition,

$$(NF)_i = \frac{(P_n)_i}{kTG_i(\Delta f)_i} \tag{8·6}$$

where G_i is the maximum power gain of the ith unit. Noise figures are frequently expressed in decibels, which must of course be converted to numbers before use in the equations to be developed. For our later purposes we need only consider the case that $(\Delta f)_1$ is so much greater than $(\Delta f)_2$ that the overall bandwidth of the system is essentially $(\Delta f)_2$. In this case, the portion of the output noise power *at the meter* which is due to the unit 1 and the noise power of the output impedance of the signal generator is

$$P_{01} = (P_n)_1 G_2 \frac{(\Delta f)_2}{(\Delta f)_1} = (NF)_1 kTG_1G_2(\Delta f)_2 \tag{8·7}$$

The portion due to unit 2 is, by definition,

$$(P_n)_2 = (NF)_2 kTG_2(\Delta f)_2 \tag{8·8}$$

Since the noise from the two sources is purely random in character, the noise powers add. However, the noise power due to the *output impedance* of unit 1 has been included in *both* P_{01} and $(P_n)_2$. Therefore the total noise power at the output meter is

$$\begin{aligned} P_n &= P_{01} + (P_n)_2 - kTG_2(\Delta f)_2 \\ &= P_{01} + [(\mathrm{NF})_2 - 1]kTG_2(\Delta f)_2 \end{aligned} \tag{8·9}$$

If we represent the properties of the overall system by symbols without subscripts,

$$P_n = (NF)kTG(\Delta f)_2 \tag{8·10}$$

where

$$G = G_1G_2 \tag{8·11}$$

then we have for the overall noise figure

$$NF = (NF)_1 + \frac{(NF)_2 - 1}{G_1} \tag{8·12}$$

It is evident that not only should $(NF)_1$ and $(NF)_2$ be made as small as possible, but G_1 should be as large as possible. In order to minimize NF the overall bandwidth should be no greater than is necessary to reproduce properly the input signals.

The overall noise figure may be measured[4] by introducing sufficient power from a signal generator (either in the form of noise or of a sinewave of the frequency at which the gain is maximum) to double the indication of the output power meter, provided, of course, the system is linear. A good approximation to an ideal signal generator is obtained by loosely coupling a real signal generator to the input terminals of unit 1 by means of an attenuator of the proper impedance. The attenuator prevents any appreciable noise power generated by the signal generator when it is turned "off" from getting into the system.

8·2 GENERAL REMARKS ON RADAR RECEIVERS[5]

The function of a receiver is to change the very small modulated radiofrequency signals delivered by the antenna to audio- or video-frequency signals of usable amplitude. A receiver thus primarily involves circuits for *amplification* and *demodulation*. The quality of a receiver is usually stated in terms of its *selectivity*, *sensitivity*, and *fidelity*.

The *selectivity* of a communications receiver is ordinarily understood to be a measure of the ability of the receiver to reject disturbances at frequencies other than the desired signal frequency. The problem of extraneous signal rejection in radar receivers is, however, somewhat different from that encountered in communications and television practice, since in the radar case interference is nearly always the result of signals occurring so close to the radar frequency that very little can be accomplished by improving the selectivity of the receiver. Thus a radar receiver should be judged more by its ability to continue to deliver intelligible information to the radar indicators in the presence of *on-frequency* interference rather than by its selectivity in the usual sense of the word.

The *sensitivity* of a radar receiver is in nearly all cases determined by its noisiness, since most receivers have sufficient amplification to give appreciable noise outputs in the absence of any

[4] Y. Beers, Radiation Laboratory Report 746, July 1945.

[5] L. W. Morrison, *Bell System Tech. J.*, **26**, 693 (1947), has given a very interesting discussion of radar receivers, including indicators.

received signal. The limitation of sensitivity by noise has been discussed in the preceding section.

The *fidelity* of a receiver determines how closely the output signals resemble the modulation envelope of the received r-f carrier. In radar receivers we are generally concerned with amplitude rather than frequency modulation, with modulation envelopes approximating rectangular pulses in shape; adequate detection of such envelopes depends more on receiver bandwidth and phase shift than on linearity of amplitude response, so that the latter consideration is relatively unimportant except as it may influence the signal-to-noise ratio in the early stages of the receiver. As a matter of fact, non-linear amplitude response in the form of limiting is customarily introduced in radar receivers.

Types of Microwave Receivers

Receivers employed in microwave radar applications are of three main types, which differ with respect to the frequency region within which most of the necessary signal amplification is achieved. *Super-regenerative* receivers employ a regenerative r-f amplifier, the oscillations in which are quenched at intervals about equal to the width of the pulses to be detected. *Superheterodyne* receivers convert the modulated r-f signal to a similarly modulated signal at a relatively low frequency, usually about 30 megacycles, before amplification. *Crystal video* receivers detect the modulation signals and amplify the resulting video signals.

The majority of microwave radar receivers are of the superheterodyne type, and we will therefore restrict attention to this type.

8·3 SUPERHETERODYNE RECEIVERS

A block diagram for a typical microwave superheterodyne receiver is given in Fig. 8·2. The modulated r-f signals from the antenna and continuous wave (*c-w*) energy from the local oscillator *LO* are mixed by the crystal mixer (first detector). The *LO* is a low power tunable oscillator such as a klystron (page 62). The resulting difference frequency is amplified by the intermediate frequency (i-f) amplifier to a level suitable for detection (demodulation) by the second detector. In systems where it is necessary to

have the main body of the receiver located at some distance from the antenna, the first two or three stages of the i-f amplifier are contained in a pre-amplifier near the antenna. The pre-amplifier output is matched, usually by a cathode follower stage (page 300), to a low impedance cable leading to the main i-f amplifier. The video output from the second detector is further amplified before distribution to the radar indicators.

In general, the chief advantage gained by heterodyning the r-f signal to a fixed intermediate frequency is that the bulk of the amplification can be achieved in a fixed-tuned amplifier, which

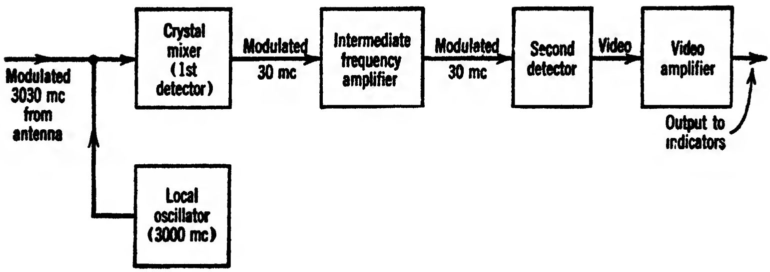


FIG. 8·2 Block diagram of a microwave superheterodyne receiver.

can be accurately lined up and then needs no further operator attention. The superheterodyne is particularly important in the microwave region because there are as yet no satisfactory low level r-f amplifiers at these frequencies. The receiver is tuned by adjusting the frequency of the local oscillator, together with occasional tuning of the frequency-sensitive parts of the r-f system. Usually the r-f tuning is broader than the *LO* tuning and therefore needs less attention.

At frequencies below the microwave region the first detector may be preceded by one or more r-f amplifier stages, and a vacuum tube mixer may replace the crystal mixer. Among the advantages gained by employing r-f amplification are improved *image frequency* rejection, improved signal-to-noise ratio, and prevention of the radiation of *LO* power. The image frequency is separated from the desired signal frequency by twice the intermediate frequency and is located on the opposite side of the *LO* frequency, so that it is received as well as the signal frequency by a superheterodyne receiver having no r-f stages. An improvement in signal-

to-noise ratio results because mixers are relatively noisy, so that it is advantageous to perform the mixing operation at a higher signal level than is possible without r-f amplification.

Crystal Mixer

The crystal mixer in most cases consists of a tungsten "cat-whisker" contacting a crystal of silicon, together with suitable connections and matching elements for introducing signal and *LO* power and removing i-f power. The use of such crystals as rectifiers for detecting moderately high level r-f signals has been discussed in Chapter 4. The tungsten-silicon contact has a non-linear current-voltage characteristic which can be expressed in the form

$$i = a_1 e + a_2 e^2 + \cdots \tag{8·13}$$

If a voltage containing two or more frequencies is impressed on the crystal, the current flowing through it will contain, in addition to other frequencies, the difference between the two input frequencies. Suppose the r-f carrier is amplitude-modulated with a single frequency, $\omega_s \ll \omega_c$, where ω_c is the carrier (angular) frequency. Then e is of the form

$$e = E_c(1 + m \cos \omega_s t) \cos \omega_c t + E_{LO} \cos \omega_{LO} t$$

if we neglect phases; ω_{LO} is the *LO* frequency. The crystal current is then

$$i = a_2 E_c E_{LO}(1 + m \cos \omega_s t) \cos (\omega_c - \omega_{LO})t + \text{additional terms} \tag{8·14}$$

The additional terms in this expression are d-c terms and terms having frequencies far removed from the i-f, except for certain terms involving $\omega_c - \omega_{LO}$ which arise from the fourth and higher terms of equation 8·13; if $a_4, \cdots$ are small, these additional terms will be small. If $\omega_c \gg (\omega_c - \omega_{LO}) \gg \omega_s$, the i-f system will be able to amplify only the first term in equation 8·14; the mixer will thus produce across an impedance in series with it an i-f voltage proportional in magnitude to the original r-f signal and modulated to the same extent. This result still holds if $\cos \omega_s t$ is replaced by a series of terms representing the Fourier expansion of a complicated modulation signal.

A typical microwave crystal mixer for use at 10 centimeters is illustrated in Fig. 8·3. The coupling loop is inserted directly into a cavity gas switch or "TR box" (page 134). *LO* power is introduced through the sidearm, by means of a variable capacitive coupling. A cup following the crystal serves as an r-f filter, the capacitance between this cup and the outer conductor constituting essentially an r-f short. It is to be noted that the mixer is fixed-tuned. Slightly better results could in principle be obtained with

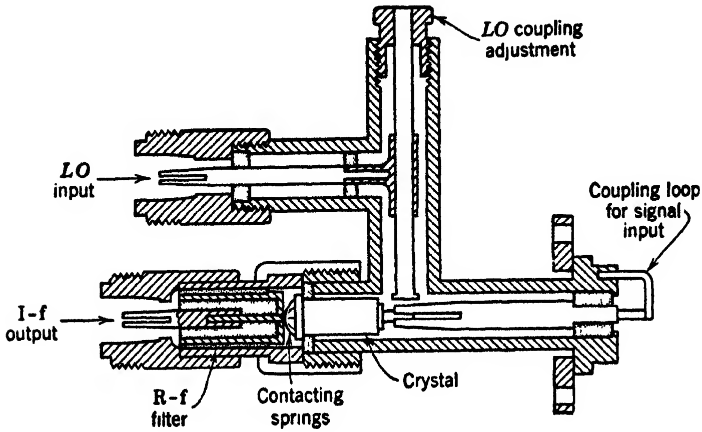


FIG. 8·3 Cross-sectional view of a fixed-tuned coaxial mixer for use at wavelengths of 9 to 11 cm. The coupling loop is designed to fit into a TR cavity.

some crystals by including a tuning device because of the variation in crystal impedances, but experience has shown that in actual practice it is better to sacrifice this small gain in order to avoid the possibility of severe loss of performance resulting from improper tuning.

Mixers for use at frequencies higher than 3000 megacycles are usually of the waveguide rather than the coaxial type. In such mixers the crystal is mounted in waveguide parallel to the electric field.

Mixer Noise

It is found experimentally that there is no significant source of noise preceding the mixer in a microwave receiver except for the

unavoidable resistive component of the antenna circuit impedance. It is therefore of prime importance to keep the mixer noise as low as possible. In general the bandwidth of the mixer is enough larger than that of the i-f amplifier so that the analysis given on page 246 for two cascaded noisy units applies. Equation 8·12 expresses the overall noise figure for such a system in terms of the individual noise figures and the gain of the first unit, in this case the mixer. The noise figure of a crystal is customarily expressed in terms of its *noise temperature*, defined by the expression

$$t_n = \frac{(P_n)_1}{kT(\Delta f)_1} = (NF)_1 G_1 \tag{8·15}$$

The gain function of a crystal which is of importance here is its *conversion gain*,

$$G_1 = \frac{\text{i-f power output from crystal}}{\text{r-f power input to crystal}} \tag{8·16}$$

Equation 8·12 can then be written in the form

$$NF = \frac{t_n + (NF)_2 - 1}{G_1} \tag{8·17}$$

where $(NF)_2$ is the noise figure of the i-f amplifier. A good crystal (type 1N21-C) for use at 10 centimeters has $G_1 = 0.28$ (-5.5 decibels) and $t_n = 1.5$.

At frequencies much above 3000 megacycles rather serious noise contributions may be made by the local oscillator itself. Noise from this source can be much decreased by using a *balanced* mixer. One form of balanced mixer, employing a waveguide *magic-tee* (page 139), is represented in Fig. 8·4; the arrows in the figure indicate the direction of the electric field in each arm. Consider the simplified case in which the *LO* output contains a term of (angular) frequency ω_n in addition to the main term of frequency ω_{LO}:

$$e_{LO} = E_{LO}(\cos \omega_{LO}t + \alpha_n \cos \omega_n t)$$

if we neglect phase differences, which are of no importance in our argument. If the crystal arms of the magic-tee are equal in length, the voltage at crystal A is, aside from phases,

$$e_A \propto E_c(1 + m \cos \omega_s t) \cos \omega_c t + E_{LO}(\cos \omega_{LO}t + \alpha_n \cos \omega_n t)$$

and that at crystal B is

$$e_B \propto -E_c(1 + m \cos \omega_s t) \cos \omega_c t + E_{LO}(\cos \omega_{LO}t + \alpha_n \cos \omega_n t)$$

Because of bandwidth restriction in the r-f circuit the voltage of frequency ω_n can appear at the crystals only if $\omega_n \approx \omega_{LO}$. The

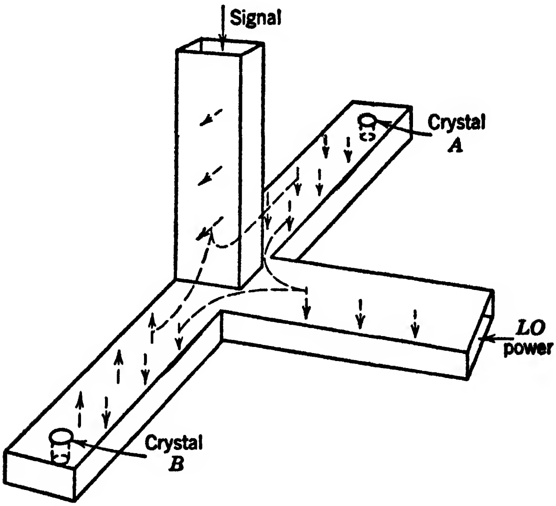


FIG. 8·4 Representation of a magic-tee balanced mixer.

currents developed in the crystals contain terms *in the i-f region* as follows:

$$i_A \propto E_c E_{LO}(1 + m \cos \omega_s t) \cos (\omega_c - \omega_{LO})t + \alpha_n E_c E_{LO}(1 + m \cos \omega_s t) \cos (\omega_c - \omega_n)t + \alpha_n E_{LO}^2 \cos (\omega_{LO} - \omega_n)t$$

$$i_B \propto -E_c E_{LO}(1 + m \cos \omega_s t) \cos (\omega_c - \omega_{LO})t - \alpha_n E_c E_{LO}(1 + m \cos \omega_s t) \cos (\omega_c - \omega_n)t + \alpha_n E_{LO}^2 \cos (\omega_{LO} - \omega_n)t$$

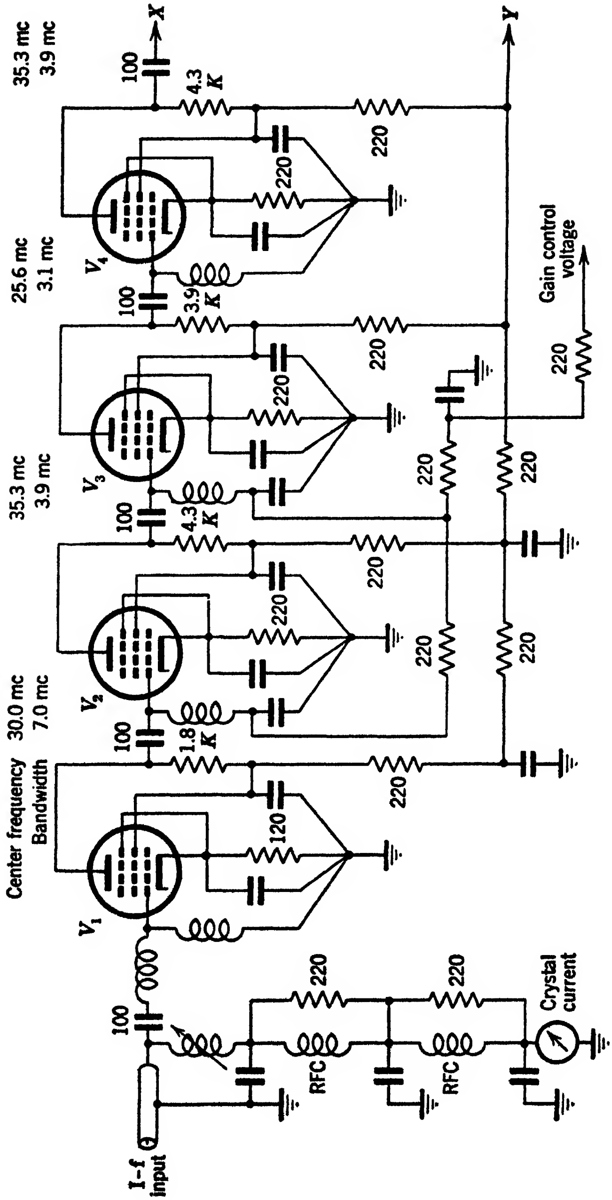
If the voltages developed across impedances in series with the crystals are applied in push-pull fashion, for example by a center-tapped input transformer, to the i-f amplifier, it is evident that the first two terms in i_A and i_B will add, while the third terms will cancel. If $\alpha_n \ll 1$, the second terms will not add significantly to the i-f output; since in general $E_{LO} \gg E_c$, the third terms might

be of appreciable amplitude in comparison with the first terms, but it is these terms which are cancelled. This reasoning applies if ω_n is replaced by a spectrum of noise frequencies.

I-F Amplifiers

I-f amplifiers are tuned amplifiers of six to twelve stages with gains of the order of 120 decibels (voltage gain 10^6). The tuned frequency is in most cases close to 30 megacycles. Thirty megacycles is a convenient intermediate frequency for the following reasons. (*a*) It is low enough so that ordinary receiving-type tubes and circuit components can be used. (*b*) It is high enough so that there is no difficulty in obtaining video signals from the second detector which are free of the i-f carrier. (*c*) Thirty megacycles is high enough so that ample bandwidth for the amplification of pulsed signals can be obtained, and coupling networks having short time constants can be employed. The latter consideration is important in securing quick recovery from the paralyzing effect of signals causing heavy overloading.

Since the gain of a crystal mixer is less than unity, much more attention has to be paid to noise developed in the input stages of the i-f amplifier than would otherwise be necessary. Until rather recently i-f amplifiers have employed 6AC7 or 6AK5 pentodes for all stages, though it was realized that a pentode is inherently more noisy than a triode with the same gain, because of partition noise (page 245). Triodes were not used because of the large input capacitance which results from plate-grid coupling through the interelectrode capacitance; this capacitance leads to poor performance at frequencies as high as 30 megacycles. A typical coupling between the mixer and the first i-f stage is illustrated in Fig. 8·5. It has been found that the signal-to-noise ratio is somewhat improved if there is a carefully chosen impedance mismatch at this point. As indicated in the figure, provision is included for measuring the d-c crystal current resulting from rectification of the *LO* signal. This current serves as a convenient indication as to whether the *LO* is actually oscillating, and as to the amount of *LO* power being coupled into the mixer. In order to have as little variability as possible in the mixer impedance at the i-f input terminals, a short length of coaxial line for connecting the i-f amplifier to the mixer is permanently fixed to the i-f chassis.



I-f input
100
RFC
220
Crystal current
V1
V2
V3
V4
Center frequency
Bandwidth
30.0 mc
7.0 mc
35.3 mc
3.9 mc
25.6 mc
3.1 mc
35.3 mc
3.9 mc
1.8 K
4.3 K
3.9 K
4.3 K
120
Gain control voltage
X
Y

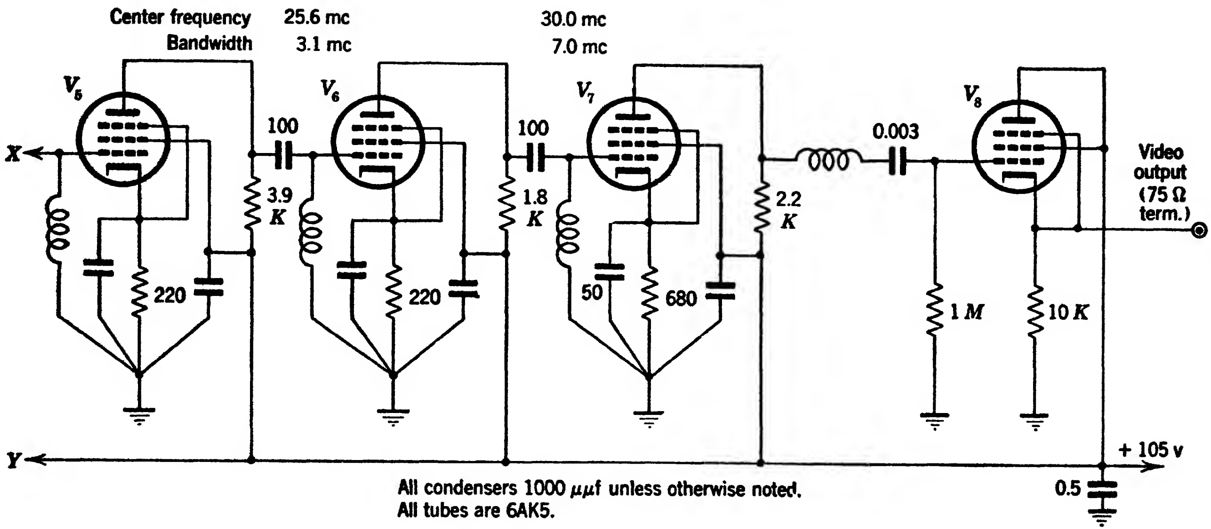


FIG. 8·5 Circuit diagram of a typical i-f amplifier. This amplifier contains two staggered triples (cf. Chap. 7) and has a bandwidth of about 8 mc.

Within the last few years it has been found possible to take advantage of the low noise characteristics of triodes without encountering difficulties because of instability due to their grid-plate capacity. One input circuit,[6] employing a triode-connected 6AK5 and a 6J6, is shown schematically in Fig. 8·6. The load impedance of the first tube, amounting to about 200 ohms, is presented by the cathode of the second tube; stability of the first tube results

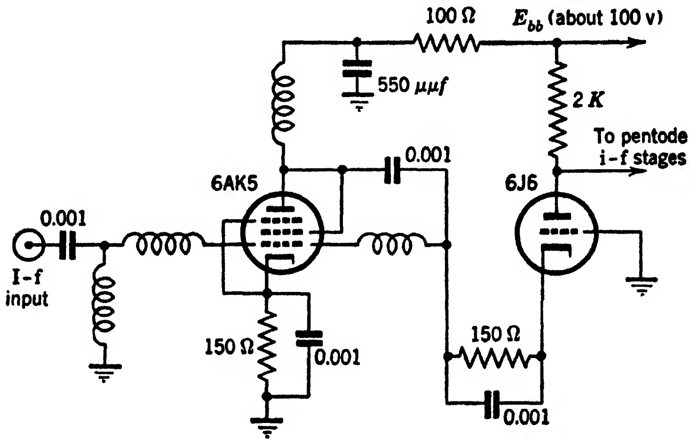


FIG. 8·6 Diagram of a grounded-cathode grounded-grid dual triode input for a 30-mc i-f amplifier. This input circuit has a noise figure about 3 db lower than a conventional pentode input.

from this low impedance. The second tube is stable in operation because its grounded grid shields the input (on the cathode) from the output. The neutralizing coil between plate and grid of the first tube confers extra stability which actually is not needed; this coil, however, is useful in helping to maintain high output impedance for the first tube, which in turn can be shown to result in making the noise contribution of the second tube very small. The gain of the combination is approximately the (triode) transconductance of the first tube times the load impedance of the second tube. This input circuit gives an i-f noise figure approximately 2 decibels lower than that ordinarily obtained with i-f amplifiers having pentode input circuits.

[6] C. P. Gadsden, Radiation Laboratory Report 699, March 1946.

Figure 8·5 shows the schematic diagram for a typical i-f amplifier,[7] consisting of two cascaded stagger-tuned triples (cf. Chapter 7) and having an overall bandwidth of the order of 8 megacycles. It is an engineering accomplishment of no mean proportions that i-f "strips" such as the one illustrated have been produced in large quantities by production line methods at surprisingly low cost per unit.

Receiver Noise Figures

I-f amplifiers employing pentode inputs have noise figures of the order of 3 to 4 decibels. With a triode input, this may be reduced to 1.5 decibels. (An i-f amplifier has been reported which, with selected tubes, gave a noise figure of 0.4 decibel!) If, in equation 8·17, we use the values of G_1 and t_n given on page 253 for a 1N21-C crystal, and $(NF)_2 = 1.5$ decibels, we obtain 8.5 decibels for the overall receiver noise figure. This is to be compared with values in the neighborhood of 16 decibels, which were considered good in 1942.

Receiver Gating

For some purposes it is necessary to have a receiver inoperative except for a selected interval during each radar repetition interval. For example, in a radar which controls the aiming of an antiaircraft gun (Chapter 11) unwanted information from targets other than the aircraft being "tracked" is prevented from reaching the tracking and computing circuits by *gating* the receiver so that it can detect a signal only during a short interval of time corresponding to the distance of the aircraft from the radar.

Receiver gating can be accomplished either in the i-f or video section. In either case one or more stages are made normally inoperative by reducing the plate or screen voltages, or by making the suppressor or control grid voltages negative or the cathode voltage positive, enough to cut the stages off. At the desired time a rectangular voltage pulse (Chapter 5) of appropriate duration and amplitude is applied to bring the tubes into the conducting region. In general an i-f amplifier is not able to amplify the frequencies contained in the gating pulse, so that the pulse does

[7] This amplifier was designed by Y. Beers of the Radiation Laboratory.

not produce any disturbance at the receiver output. On the other hand, if a video amplifier is gated, the gate pulse will produce a so-called pedestal at the output on which the amplified signals will ride. This pedestal may be desired under some conditions; if the gated signals must be free of any pedestal, compensation may be arranged. This is done by having the plate current, which is carried during the gate by a gated tube, carried between the gates by a tube with no signals on its grid connected in parallel with the gated tube and itself turned off by an inverse gate pulse when the gated tube is turned on. A circuit is given in Fig. 10·20 which accomplishes pedestal-free video gating.

Receiver Gain Control

It is important to be able to control the overall gain of an i-f amplifier to compensate for variations in transconductance of tubes, as well as for other purposes. If the gain of two or three stages increased by small amounts as a result of tube changes, one could well reach a condition where even noise signals would saturate the later stages in the amplifier. The most satisfactory method of gain control is to increase the negative bias on one or more of the i-f grids. In Fig. 8·5 the second and third stages are gain-controlled. Gain control should not in general be applied to a stage which contributes appreciable noise, as shown in the discussion in Section 8·1. Completely cutting off a single stage, or even removing the tube altogether, decreases the gain only 30 to 40 decibels because of direct capacitive or inductive feed-through in the wiring.

Automatic Gain Control

If an individual signal is selected by receiver gating (page 259), the video output resulting from this signal can be used to develop a d-c voltage which can then be fed back to one or more i-f grids to give automatic gain control (AGC). As usual with feedback systems, care must be exercised not to allow phase shifts which will cause regeneration at any frequency amplified by the closed loop. Perhaps the simplest means for developing the AGC voltage is the "integration" circuit shown in Fig. 8·7. The time constant RC should be large compared to the repetition interval but small compared to the period of the expected variations in the

amplitude of the video signals. The output of the circuit of Fig. 8·7 is inverted by an amplifier stage and fed to the i-f grids through a cathode follower (page 299).

An interesting application of AGC is in the detection of small amounts of audio amplitude modulation of radar signals such as may result from reflections from the rotating propeller of an airplane.[8] Obviously it is necessary that saturation at any point in the receiver be carefully avoided so that the modulation will not be removed; in this case one must apply gain control to four or five i-f stages if large signals are to be handled. Circuits for de-

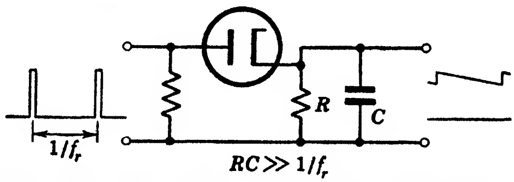


FIG. 8·7 Basic circuit of a diode integrator for developing a gain control voltage.

tecting such modulation are described on page 309. The detected output is a good source for the necessary AGC voltage, if it is passed through a low pass RC filter to remove the modulation frequency in order to prevent degeneration of the modulation signal. It is interesting to observe that in this application the lower limit to the amount of audio modulation which can be detected on strong signals will usually be placed by the noise arising in stages late in the i-f amplifier the gain of which is uncontrolled, since the noise in the receiver output from these stages is not reduced as the receiver gain is reduced. It has been found that under favorable conditions audio modulation as small as 0.01 per cent can be detected.

Second Detector

The second detector in most microwave receivers is of the diode type, operated as a *linear* [9] detector; that is, the applied i-f signals

[8] R. M. Ashby, F. W. Martin and J. L. Lawson, Radiation Laboratory Report 914, March 1946. See also J. M. Sturtevant, Radiation Laboratory Report 654, Jan. 1945.

[9] For detailed discussions of detectors the reader is referred to standard textbooks.

are large enough so that their peaks are above the non-linear portion of the diode characteristic even for signals modulated only by noise. In this way one avoids discrimination against signals smaller than noise, which may actually be distinguishable from noise because of their shape The diode thus functions simply as a rectifier. It is followed by a low pass filter to keep i-f signals out of the video amplifier. The detector output is usually made negative so that limiting can be accomplished in the first video stage.

Plate detection is sometimes employed, as in the circuit of Fig. 8·5, a pentode (V_7 in the figure) operated near cutoff, usually self-biased, being used. A sharp cutoff tube is used so that small signals reach the linear portion of the characteristic.

8·4 ACCESSORIES FOR SUPERHETERODYNE RADAR RECEIVERS

Several refinements and additions have been developed for microwave radar receivers which either increase their general performance and reliability, or are needed to adapt them to particular functions. Some of these are discussed in the following paragraphs.

Automatic Frequency Control (AFC)

For proper performance the *LO* frequency must be held at a value separated from the transmitter frequency by the intermediate frequency, with an accuracy determined by the i-f bandwidth. There are several factors which frequently make it desirable that the *LO* frequency be automatically controlled. Very high frequency oscillators, both low and high power, show considerable tendency to drift in frequency during prolonged warm-up periods. The frequency of a transmitter, such as a magnetron, is "pulled" by changes in the impedance into which the transmitter works (cf. Chapter 3); an impedance change may take place, for example, during antenna rotation if the r-f line includes rotating joints or if the antenna is housed in a "radome" from which there are varying reflections. A 10-centimeter magnetron may be "pulled" as much as 8 megacycles by an impedance mismatch causing a voltage standing wave ratio of 1.5; in the case of a 3-centimeter

magnetron this figure is about 25 megacycles. In general, the average performance of a radar is apt to be improved by relieving the operator of all possible adjustment responsibilities, particu-

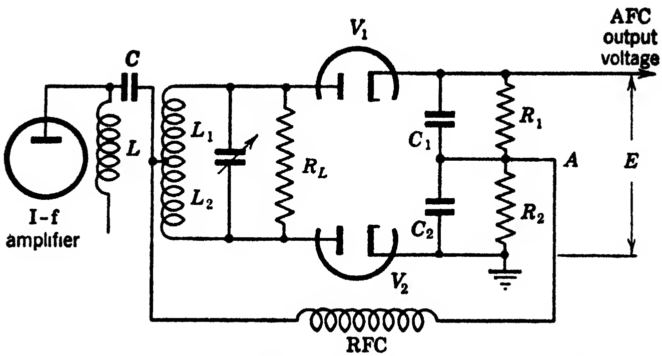


FIG. 8·8 Diagram of an i-f discriminator for obtaining a voltage for automatic control of the frequency of a local oscillator.

larly under conditions where his attention is largely absorbed by other duties.

AFC systems depend in nearly all cases on a frequency discriminator, operating at the intermediate frequency, similar to those used in frequency-modulation receivers. A typical discriminator is shown in Fig. 8·8. L and L_1, L_2 comprise a doubly tuned

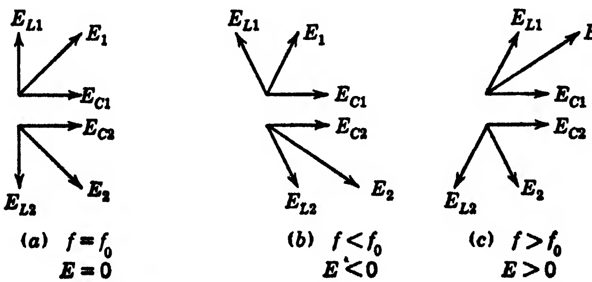


FIG. 8·9 Vector representation of the voltages developed by the discriminator of Fig. 8·8.

transformer. There is capacitive coupling, supplied by C, between the primary and the secondary, as well as inductive coupling. The voltages developed at the diode plates due to the capacitive coupling (E_{C1} and E_{C2}, Fig. 8·9) are in phase at all frequencies.

At $f = f_0$, the voltages due to the inductive coupling (E_{L1} and E_{L2}) are 180° out of phase with each other and 90° out of phase with the capacitive voltages; if $f < f_0$ the phase relations are as shown in Fig. 8·9 (*b*), and if $f > f_0$ they are as shown in Fig. 8·9 (*c*). The secondary voltages are detected by the diodes and "integrated" by R_1, C_1 and R_2, C_2. If the time constants R_1C_1 and R_2C_2 are sufficiently large, we need consider only the magnitudes $|E_1|$ and $|E_2|$. It is seen that the d-c output voltage E vanishes at $f = f_0$, is negative for $f < f_0$, and is positive for $f > f_0$. The discriminator characteristic is of the form illustrated in Fig. 8·10.

The discriminator is essentially a balanced detector; it is therefore possible to take video signals, or an automatic gain control

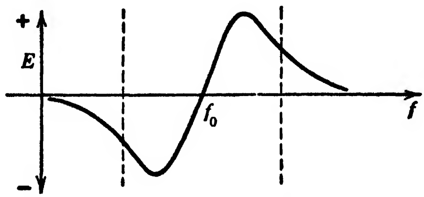


FIG. 8·10 Output of the discriminator of Fig. 8·8 plotted as a function of frequency.

voltage (see below) from point A in Fig. 8·8. It might be supposed that the discriminator would follow the main i-f amplifier and serve as both discriminator and detector. This is very seldom the case in radar practice, since the AFC could then be actuated by strong echoes or by strong interference pulses. It is much better to take advantage of the fact that high level transmitter output pulses are available for AFC purposes. A separate AFC mixer is usually coupled loosely into the r-f line, the *LO* power being supplied by the receiver *LO*. Since high level pulses are available, a low gain broadband i-f amplifier will be sufficient to drive the discriminator.

The discriminator output can be used in various ways. If the *LO* frequency can be changed by varying the voltage on one of its electrodes (such as the reflector of a McNally tube; see page 70), the discriminator output can be amplified by a d-c amplifier and applied in the proper phase to this electrode. If a tuning knob on the *LO* must be turned in order to vary its frequency, the discriminator output can be applied as an error signal to a small

servomechanism (cf. Chapter 9). In any case, the "tightness" of the AFC will depend on the gain of the feedback loop; care must be taken that oscillations are not produced as the result of phase shift in the feedback loop.

A quantity of importance in an AFC system is the "pull-in" range, that is, how far the *LO* frequency can be removed from its proper value and still have the AFC able to resume control. If the discriminator output is used directly through a feedback loop as mentioned in the preceding paragraph, the region of control is about as indicated by the dotted lines in Fig. 8·10. With a re-

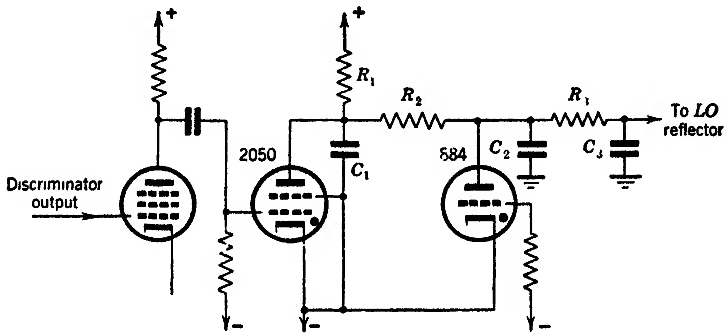


FIG. 8·11 Schematic diagram of a sweep circuit for increasing the "pull-in" range of an automatic frequency control system.

ceiver having a 2-megacycle bandwidth, the discriminator peaks are adjusted (by means of the loading resistor R_L) to occur at about $f_0 \pm 2$ megacycles, so that the pull-in range would be approximately $f_0 \pm 4$ megacycles. The pull-in range can be greatly increased by having the discriminator output control a frequency-sweeping circuit. Such a circuit, devised at the Radiation Laboratory, is illustrated schematically in Fig. 8·11. The 884 thyratron serves as a relaxation oscillator, its grid bias being adjusted so that the voltage across the condenser C_2 sweeps through the complete tuning range of the *LO*. The period of the relaxation oscillations depends chiefly on the time constant R_2C_2, which is much larger than R_1C_1. The 2050 thyratron stops the sweep of the 884 at the proper point in the following way. Suppose the *LO* is tuned on the high frequency side of the transmitter. During the charging of C_2 toward more positive voltages, the *LO* frequency is decreased, so that the intermediate frequency is also

decreased. As long as the intermediate frequency is greater than the balance frequency of the discriminator, positive output pulses are obtained from the discriminator, and therefore negative pulses which have no effect are impressed on the grid of the 2050. As soon as the intermediate frequency becomes less than f_0, output pulses are applied to the grid of the 2050 and cause this tube to fire, thus reducing the voltage across C_1 and, much more slowly, that across C_2. This action tends to increase the intermediate frequency, and soon the 2050 stops firing, and the voltage across C_2 again starts to rise. Thus the voltage across C_2 oscillates with small amplitude about the proper value, and the *LO* is held at very nearly the correct frequency. This swept AFC will pull in if the proper frequency is within the electrical tuning range of the *LO*.

Frequency Stabilization

The AFC using a discriminator is essentially a frequency stabilization system based on the resonant frequency of the discriminator as the reference point for the stabilization. It is obviously advantageous to perform the stabilization at the relatively low intermediate frequency. However, there are cases in which this cannot be done. For example, a receiver for detecting signals transmitted from a physically remote transmitter, such as a beacon (cf. Chapter 11) or a communications or relay (cf. Chapter 12) receiver, should be stabilized in frequency, but it may be unsafe to perform this stabilization at the intermediate frequency because of the danger of locking on interfering signals. The only alternative is to perform the stabilization at the *LO* frequency. This may be done in the microwave region by using an r-f cavity as the comparison standard. Three methods for accomplishing this are briefly described in the following paragraphs.

A low power oscillator tube may be stabilized by being more or less tightly coupled to a high Q cavity. This method has the drawback that it is frequently found that an unstable equilibrium results, and that the tube stops oscillating if its frequency drifts slightly.

An ingenious method for obtaining stabilization information from a cavity has been devised as the result of a suggestion made by J. Halpern of the Radiation Laboratory. The cavity frequency is modulated by moving the end plate slightly by means of a loud speaker driving mechanism. If the modulating frequency is 60

cycles, and energy is fed into the cavity, the output is amplitude modulated with a fundamental of 120 cycles if the input frequency equals the resonant frequency of the cavity. If the input frequency differs from the resonant frequency, the modulation of the output has a fundamental of 60 cycles, and the phase of the fundamental shifts 180° as the input frequency goes through the resonant

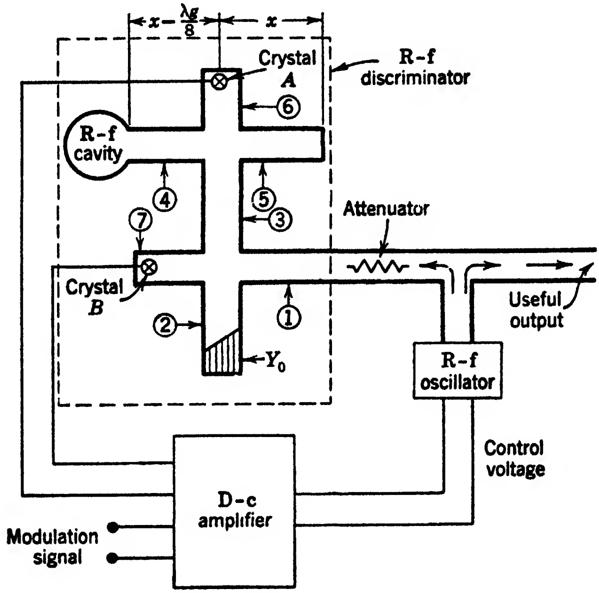


FIG. 8·12 Schematic diagram of Pound's r-f discriminator, showing its application in the stabilization of an r-f oscillator.

value. Thus, if the detected cavity output is filtered to remove all frequencies above 60 cycles and is then amplified and applied to a phase-sensitive device such as a two-phase motor or a phase-sensitive detector, a control mechanism can be obtained.

A very elegant frequency stabilization method, developed by Pound,[10] makes use of the fact that the impedance of a cavity depends on the frequency. A schematic representation of the r-f discriminator employed is given in Fig. 8·12. Power from the oscillator to be stabilized is fed to a magic-tee (see page 139) where it is equally divided into arms 2 and 3. The power going into

[10] R. V. Pound, Radiation Laboratory Reports 662, 815, and 837, 1945; *Rev. Sci. Instruments*, **17**, 490 (1946); *Proc. I.R.E.*, **35**, 1405 (1947).

arm 2 is absorbed in a matched termination, while that in arm 3 is divided by a second magic-tee into arms 4 and 5. Arm 5 is shorted, so that the power in that arm is reflected. Arm 4 is terminated in an r-f cavity, and is shorter than arm 5 by one-eighth of the guide wavelength, λ_g. If the frequency f is much smaller or larger than the resonant frequency f_0 of the cavity, the cavity appears as a short circuit, so that the power reflected from it is $\pi/2$ out of phase with the power reflected in arm 5. It is a property of the magic-tee that this leads to excitation of waves of equal amplitude in arms 3 and 6. The power in arm 6 is absorbed by crystal A, while the power in arm 3 is again divided by the first magic-tee, so that half of it is absorbed by crystal B. Suitable adjustment in the d-c amplifier is provided so that the output of the amplifier is zero when the output of crystal A is twice that of crystal B. At $f = f_0$, the cavity gives a reflection of the same or opposite phase to that from a short circuit, so that again waves of equal amplitude are excited in arms 3 and 6. For $f > f_0$, the cavity presents a capacitive reactance, and the wave excited in arm 6 is greater than that in arm 3; for $f < f_0$ this situation is reversed. Thus a discriminator characteristic similar to that shown in Fig. 8·10 is obtained. The characteristic is reversed by making arm 4 longer by $\lambda_g/8$ than arm 5.

As indicated in Fig. 8·12, either voice or video modulation can be superimposed on the control voltage which is fed to the control electrode (usually the reflector) of the r-f oscillator. The control amplifier must be made insensitive to the modulation frequencies since the modulation would otherwise be strongly degenerated.

With a carefully designed amplifier, Pound found it is possible at 10,000 megacycles to obtain short time stability of the order of 1 part in 10^8, and long time stability of 1 part in 10^6, without employing any special precautions such as temperature-independent cavities.

It is interesting to point out that with the cavities at present available at 10,000 megacycles, a movement of 1 Ångstrom (10^{-8} centimeter) of the end plate changes the resonant frequency about 10 cycles. This sensitivity could be considerably increased for ultramicrometer applications by proper design of the cavity.

Anti-Jamming and Anti-Clutter Circuits

Under wartime conditions much attention had to be paid to protecting a radar set from attempts on the part of the enemy to

render it useless, that is, to "jam" it. Anti-jamming considerations are also important under normal conditions, since it turns out that circuits which are successful in partially overcoming intentional jamming are also effective in minimizing difficulties arising from jamming due to other causes. The subject of jamming and protection against jamming is very extensive, as evidenced by the fact that there was a large laboratory in this country during the war which was concerned solely with radar countermeasures; we will be able to do no more here than mention briefly a few of the more important circuits which have been developed for anti-jamming purposes. Our discussion is based on that given by Lawson, *et al.*[11]

Intentional jamming may consist of the use of reflecting material such as "window" (strips of metallic foil which cause numerous fluctuating echoes and thus obscure aircraft in the vicinity), or of radio waves, either unmodulated or modulated. In the latter case, the two commonest forms of modulation are more or less long pulses (termed "railings" because of the appearance, on an A-scope, of pulses which are not synchronized with the radar), or noise. Accidental jamming may result from strong echoes from land targets ("land clutter"), from rough water surfaces ("sea clutter"), or from clouds, or it may be due to interference from other high frequency equipments, in which case it will usually take the form of c-w or railings jamming.

Circuits for protection against jamming and clutter accomplish no more than a good manual gain control; they are important because it would be quite impossible to follow with a manual control the rapid changes in jamming and clutter taking place with range variations and antenna rotation. The chief aim of such circuits is to prevent saturation of the receiver and consequent loss of signal visibility. A signal of power 70 decibels above noise power in 40 decibels clutter would look about like a 30-decibel signal if the receiver gain were appropriately reduced, but it would be invisible with normal gain since the clutter would saturate the receiving system, particularly if an intensity-modulated indicator such as a PPI (Chapter 6) were used. Such indicators have a small dynamic range, of the order of 10 to 20 decibels, so that

[11] V. Josephson, L. Linford, J. Lawson, and C. Palmer, Radiation Laboratory Report S-52, August 1945. See also P. R. Bell and F. M. Ashbrook, Radiation Laboratory Report S-8, Feb. 1944.

signals must be correspondingly limited by a video limiter stage.

Four types of circuits have proved themselves useful in giving protection against jamming and clutter:

(*a*) *Sensitivity time control* (STC) is a circuit which controls the gain of the receiver as a function of the time after the initial radar pulse. Usually the gain is reduced at the time of the initial pulse and is then gradually increased to the normal value according to a curve determined by the circuit constants. Since this circuit makes the gain a function of range, it is in general useful only against interference having radial symmetry and decreasing with increasing range. Such interference is frequently caused by sea clutter, or in some cases by land clutter.

(*b*) *Instantaneous automatic gain control* (IAGC) rapidly decreases the gain of an i-f stage when the output of that stage increases beyond a value determined by the circuit constants, and thus serves to prevent saturation of that stage. It is usually advisable to protect the last two or three stages with IAGC. The time constant of the operation of the circuit is of the order of 20 microseconds.

(*c*) *Fast time constant* (FTC) is a coupling between the second detector and the first video stage having a short time constant which serves to remove or attenuate, as a result of differentiating action, d-c and low frequency terms encountered in c-w or low-frequency modulated c-w jamming. The short time constant is usually of the order of the radar pulse width.

(*d*) *Detector balanced bias* (DBB) automatically supplies a bias to the second detector which is sufficient to prevent the high-frequency components of noise-modulated c-w jamming or clutter from saturating the video section of the receiver. Since the circuit can be made very fast-acting it is necessary to insert a time delay so that discrete signals will not be reduced in amplitude.

FTC and DBB circuits are most effective when used in conjunction with IAGC. The FTC and IAGC combination is effective for jamming by unmodulated c-w or c-w modulated at frequencies below the FTC cutoff frequency, while the DBB and IAGC combination is best in cases of noise-modulated jamming and most types of clutter, especially clouds. In other cases, the most effective combination must be determined by trial and error.

Very effective protection against on-frequency jamming produced by an interfering transmitter is obtained by tuning the

radar transmitter and receiver to a different frequency. The development of broadband r-f components and tunable magnetrons has made possible the construction of radar systems which can be tuned to a new frequency by a single control knob.[12] Such a system may also be useful in reducing land clutter in localized regions of the system's coverage since it has been found that many land echoes show considerable frequency sensitivity.

High power pulsed radar systems operating in close proximity may suffer from mutual interference even when the radars are not tuned to the same frequency, because the receiving system of a radar does not sufficiently attenuate extremely strong off-frequency signals. Interference in such cases may also arise from pickup of the large modulator video pulses of a nearby system. Considerable success in the removal of interference of this type has been achieved by blanking the receiver of the set suffering interference by means of a gating pulse applied to one or more i-f grids and synchronized with the transmitted pulse of the interfering set. The blanking pulse may be developed by means of a special receiver tuned to the other system, or by direct cabling of a trigger pulse from one system to the other. Obviously such a procedure chops holes in the coverage of the protected system, but this is far better than to have its indicators covered by interfering signals. In some cases one radar can be completely synchronized with another, and thus the interference avoided altogether.

PROBLEMS

8·1 Show that the maximum power transfer from a generator to a load takes place when the load impedance is the complex conjugate of the generator impedance, and calculate what fraction of the total power is transferred to the load.

8·2 Show that for a single-tuned filter (Section 7·1) the noise bandwidth as defined in equation 8·5 is $\pi/2$ times the half-power bandwidth.

8·3 Draw a complete block diagram for a radar receiver having automatic gain control in which the gain control voltage is developed from the receiver output during a 10-μsec period which can be made to occur at any desired time during the interpulse period.

[12] J. E. Cook and J. E. Richardson, Radiation Laboratory Report 911, March 1946.

CHAPTER 9

SERVOMECHANISMS AND COMPUTERS

The problem of precise remote control is one of increasing importance. For example, a significant part of the cost of a chain-reacting plutonium pile is in the remote control of the neutron-absorbing rods and in the remote indication of many factors, and the successful development of such piles is due in no small measure to the fact that the necessary remote controls and indicators could be devised. During the war this problem was urgently faced in connection with gun direction, and in the positioning of the yokes of radar indicators (Chapter 6). Also, a great deal of the technique for remote control either involves electrical computation or can in turn be applied to computing if desired.

Devices which are capable of giving precise remote control are in nearly all cases "servomechanisms." We here describe some simple forms of servomechanisms and suggest some elementary methods of electrical computing.

Servomechanisms may be briefly described as closed-cycle control systems. The meaning of this term will be made clearer below. Many of the automatic control devices used for scientific and industrial purposes can properly be classed as servomechanisms since they involve feedback loops.

The theory of servomechanisms is rather complex and will not be gone into here. The reader is referred to other sources [1] for

[1] *Radiation Laboratory Technical Series;* E. S. Smith, *Automatic Control Engineering,* McGraw-Hill Book Co., 1944; L. A. MacColl, *Fundamental Theory of Servomechanisms,* D. Van Nostrand, 1945; R. E. Graham, *Bell System Tech. J.*, **25**, 616 (1946).

theoretical discussions, as well as for more detailed consideration of practical applications.

The term computers includes a wide variety of electrical and mechanical systems which facilitate mathematical computations. Considerable publicity has been given to the very complicated computing machines which have been invented during the last ten years or so for handling computational problems which would be extremely tedious, or even impossible of human solution, if attacked by older methods. Wartime applications resulted in the development of many other devices which carry out more or less involved computations automatically. For example, data obtained by an optical or radar gunsight concerning the position and rate of change of position of an aircraft target, together with the other factors entering into ballistic calculations, can be fed into a computer which will then automatically arrive at the proper aiming of an anti-aircraft gun.

Computers may be divided into two classes.[2] *Digital* computers are machines which, like ordinary calculating machines, operate on numbers expressed in digit form, in either the decimal or some other system. *Analogue* computers, on the other hand, operate on numbers represented by electrical quantities, such as voltage or resistance, or mechanical quantities, such as the rotational position of a shaft. In general, digital computers are capable of a higher order of precision than are analogue computers; on the other hand, the latter are better adapted to situations where an essentially continuous indication or application of the results of the calculation is required. We shall concern ourselves in this chapter only with analogue computers involving electrical analogues and shall restrict ourselves to a few simple examples in an attempt to illustrate the kind of operations which such computers are capable of carrying out.

9·1 SYNCHROS

An important application of servomechanisms is in causing a loaded shaft to turn in synchronism with another shaft. Nearly all systems for synchronizing the rotation of two shafts are based on

[2] See, for example, G. R. Stibitz, Applied Mathematics Panel, NDRC Report 171.1R, Feb. 1945.

various types of *rotary transformers* known as *synchros.*[3] If two closely spaced coils are oriented with their axes parallel, an alternating current in one will induce an alternating voltage across the other; if one coil is rotated with respect to the other the induced voltage will vary (approximately) as the cosine of the angle between the coils. Such a device constitutes a rotary transformer. With proper design and a constant exciting voltage across one coil, the voltage across the other coil can be used as an accurate

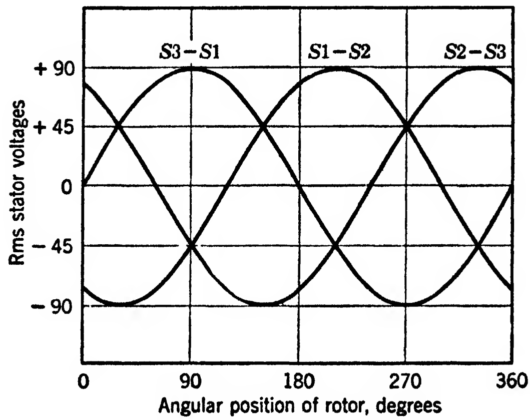


FIG. 9·1 Stator voltages of a synchro generator as functions of the angular position of the rotor. Voltages represented as positive are in phase with the input to terminals $R1$ and $R2$.

indication of the angle between the coil axes. However, the output voltage will be zero for both 90° and 270°. To remove this ambiguity one of the coils can be replaced by three coils oriented 120° apart in space. A *synchro generator* (G) is such a device, having a salient pole rotor with one winding and a so-called three-phase[4] stator with three windings either Δ- or Y-connected. When an alternating excitation is applied to the rotor, the *amplitudes* of the three output voltages vary as shown in Fig. 9·1. $S1$, $S2$, $S3$ are the conventional representations of the three stator

[3] Various trade names, such as Selsyn and Autosyn, are sometimes used in place of the generic term synchro.

[4] It is to be noted that the voltages across the three stator windings differ in phase from the excitation voltage by either 0° or 180°; the *amplitudes* of these voltages differ from each other by 120° or 240° in phase as the rotor is turned at a constant rate.

terminals, and $R1$, $R2$ of the rotor terminals. The instantaneous voltages across the rotor and one of the stator windings are shown in Fig. 9·2, for the case of 60-cycle excitation and the rotor turning at 150 revolutions per minute. It is evident that each value of the rotor position between 0° and 360° gives a unique set of stator voltages, so that all ambiguity has been removed.

The most usual indicating device for interpreting the stator voltages is a *synchro motor* (F) or *follower*. This is essentially the

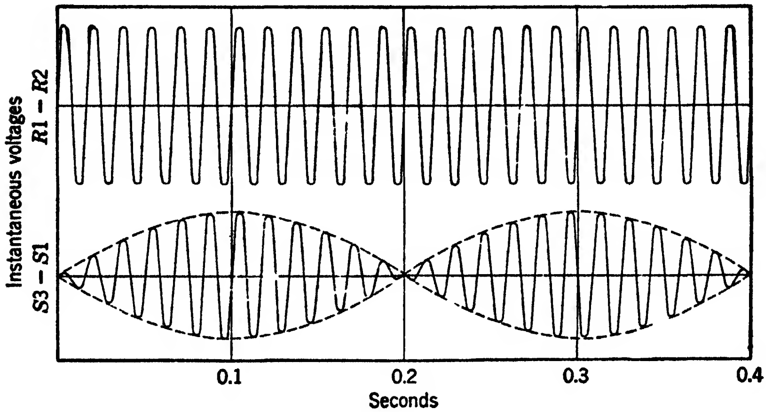


FIG. 9·2 Representative instantaneous voltages across rotor and stator terminals of a synchro generator as the rotor is turned at 150 rpm.

same as the generator, except that it includes a mechanical oscillation damper. A generator and motor are connected as illustrated in Fig. 9·3 to form a simple data transmission system.

Since no torque is developed by the motor when its rotor position is the same as that of the generator, a small error which increases with increasing motor load is unavoidable. For this reason a simple system such as shown in Fig. 9·3 is usually used only with very light loads. The error in the load position can be reduced by gearing both the synchros to run at some multiple of the input and output shaft speeds, although this procedure obviously reintroduces ambiguity in the angle information. A simple method for handling this situation is described on page 213. The Navy type $5F$ synchro motor when used with a $5G$ generator has a no-load error of approximately ±0.6°. The rotation of the motor can be reversed by interchanging any pair of stator leads

on either the motor or the generator. By convention, positive angles correspond to clockwise rotation of the shaft when the synchro is viewed from the shaft end. One generator may control several motors, particularly if the generator is of larger size than the motors.

Two other types of synchro units are important. A *control transformer* (*CT*) has the usual three stator windings and a non-

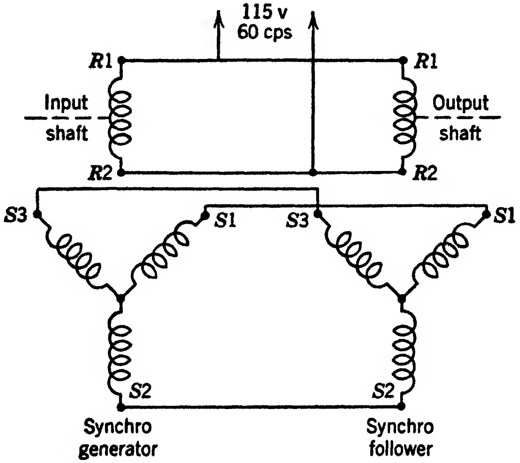


Fig. 9·3 A simple data transmission system composed of a synchro generator and a synchro follower.

salient pole rotor with one winding. If the stator windings are connected to those of a generator, the voltage across the rotor varies sinusoidally in amplitude with the relative position of the two rotors, undergoing a phase change of 180° at 0° and 180°. This rotor signal is called an error signal, and may be utilized in a servomechanism as described in the next section. The *CT* rotor is designed to drive a high impedance load so that very little current flows through it. Because of this and the non-salient pole construction of the rotor, very little torque is developed by a *CT*. A *CT* constitutes an inductive load, so that, unless an oversize generator is used or if several *CT*'s are driven by one generator, power factor correction is advisable. This can be accomplished by a triple condenser in Δ- or Y-connection across the stator leads.

A *differential generator* (*DG*) is a device by which an angular displacement can be inserted between a generator and a motor or

control transformer. Both the stator and rotor of a *DG* have three windings. When connected as shown in Fig. 9·4, the angular position taken by the motor is the sum of the angular positions of the *G* and the *DG*. If a *CT* is used with a *DG*, power factor correction may be necessary as indicated above; the condenser may be connected on either side of the *DG*.

Conventions have been established as to the zero position of the rotors of synchros. Naturally these are related to the voltages

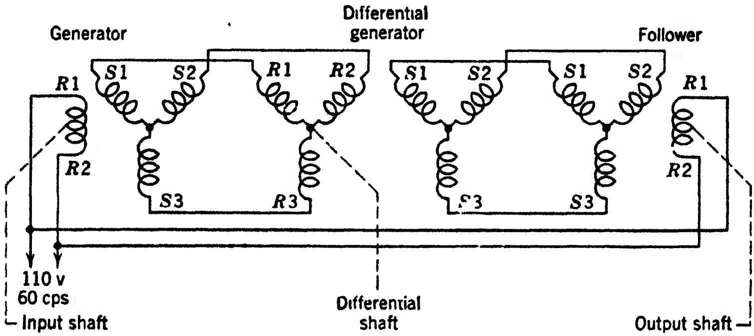


FIG. 9·4 Diagram illustrating the connection of a synchro differential generator between a synchro generator and follower. The motion of the output shaft can be either the sum or difference of the motions of the input and differential shafts.

developed at various terminals. The conventions are such that the devices may be set in their zero positions in the following ways. (*a*) To zero a generator or motor, an excitation voltage of about 80 volts (for a 110-volt instrument) is applied to the rotor. *S*2 is connected to *R*1, and the rotor is turned (or the stator is turned, if the setup is such that the rotor cannot be turned) until the voltages between *S*1 and *R*2 and between *S*3 and *R*2 are both about 37 volts. The setting is then refined by disconnecting *S*2 from *R*1 and adjusting until the voltage between *S*1 and *S*3 is zero, with full excitation on the rotor. (*b*) A control transformer is set to zero by applying about 80 volts between *S*1 and *S*3, connecting *R*2 and *S*3, and leaving *S*2 unconnected. At the zero position the voltage between *R*1 and *S*1 is a minimum and equal to about 35 volts. The setting is refined by applying about 80 volts between *S*2 and *S*3, with *S*1 connected to *S*3, and adjusting to make the voltage across the rotor vanish.

Simple synchro data transmission systems in which a generator synchro drives a motor are satisfactory for light loads in cases where the highest precision is not necessary. For heavy loads, such as gun turrets or radar antennas, and in cases where the angular error must be very small, torque amplification is necessary. This is achieved by a *servomechanism.*

It is difficult to give an inclusive definition of a servomechanism. Most servos follow the basic block diagram given in Fig. 9·5. An

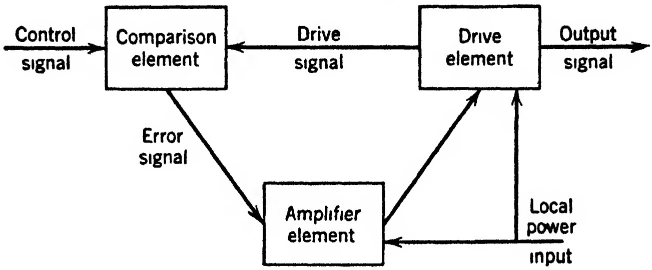


FIG. 9·5 Basic diagram of a servomechanism.

essential feature of a servomechanism is the feedback loop which feeds a drive signal, usually identical with the output signal, to a comparison element. The comparison element compares the control and drive signals, and from this comparison derives an error signal, which is in many cases simply the difference between the control and drive signals. The error signal is the effective cause which operates the output system composed of the amplifier and drive elements. It is an essential characteristic of a servomechanism that the error signal supplies only a negligible fraction of the power required to actuate the output system; there is always a local source which supplies practically all the useful output power and the power lost by dissipation. It is seen, then, that a servomechanism is essentially a *feedback amplifier*; as would be expected, the problems involved in securing satisfactory, stable operation of a servomechanism can be handled by methods similar to those evolved in treating feedback amplifiers. These problems, even in the simplest cases, can be solved in detail only by the application of rather complicated theory and computations. Fortunately, in simple cases satisfactory, though probably not optimum,

results can frequently be obtained by more or less educated trial and error methods.

As an example of a servomechanism we may consider the familiar thermostat-relay-heater temperature regulator. The control signal is the setting applied to the thermostat, which serves as the comparison element. If the temperature of the body under control differs from the setting of the thermostat, an error signal is supplied to the relay which functions as the amplifier element and turns on or off the heating (drive) element. In this case the drive and output signals are the same, namely the temperature of the body under control.

This system has a deceptively simple appearance. It is evident that many characteristics of the system, such as the heat capacity and thermal conductivity of the heat transfer medium, the effectiveness of stirring, and the rate of heat input, will affect the functioning of the system in important measure. Furthermore it is a highly non-linear system, the theory of which is in general much more difficult than that of linear servo systems. In spite of this complexity, trial and error methods have brought this type of system to a degree of development satisfactory even for many rather exacting applications.

9·3 SERVOMECHANISMS FOR THE CONTROL OF SHAFT ROTATIONS

An important type of servo system is used to cause an output shaft to rotate in synchronism with an input shaft. Figure 9·6 gives the block diagram of such a system. In this case, the control signal is the three-wire synchro data from a synchro generator, the comparison element is a control transformer, and the error signal is the alternating voltage taken from the rotor winding of the *CT*. The error signal is amplified by a vacuum tube amplifier, and the amplified output energizes a drive motor. The rotation of the motor shaft is transmitted through suitable gearing back to the shaft of the *CT* in such a fashion as to tend to reduce the error signal to zero. To remove the ambiguity inherent in the sine function nature of the error signal, provision must be made in either the amplifier element or the drive motor (in the latter in the diagram of Fig. 9·6) for comparison of the phase of the error signal

and that of the alternating voltage used to excite the synchro generator. A zero error signal will still result from an angle error of

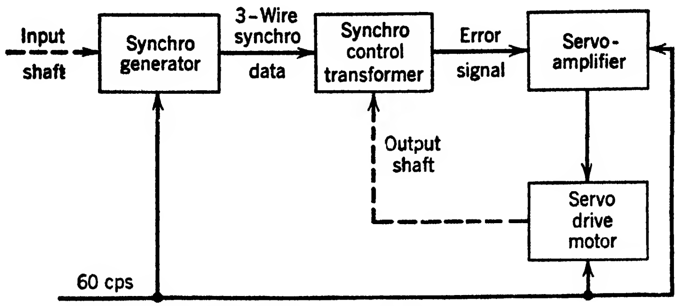


FIG. 9·6 Simplified block diagram of a servomechanism for controlling a shaft rotation. Dotted lines represent mechanical connections.

180° as well as from one of 0°, but this is a position of unstable equilibrium and hence gives no trouble in practice.

A Low Power Servomechanism

The circuit diagram of a typical low power servo system is shown in Fig. 9·7. This system is suitable for driving a light load, such as a PPI deflection yoke, with a very small error. The error signal from the rotor of a synchro *CT* is split by a transformer into two channels one of which includes a twin-tee filter (see below). The signals are applied to the grids of a cathode-coupled 6SL7 amplifier (page 300) which mixes them and gives a balanced push-pull output to drive a pair of 6L6 power amplifiers. The latter are transformer-coupled to one phase of a two-phase reversible induction motor, the other phase of which is excited from the same 110-volt line as is used for synchro excitation. A condenser in series with this winding of the motor produces a 90° phase shift which is essential to proper operation of the motor. Depending on the transformers used, some phase correction of the amplifier output may also be necessary. Since the control motor is a phase-sensitive device, the direction of its rotation is reversed by reversal of the phase of the error signal. It is important that the motor and its load have low inertias in order to avoid "hunting." The inertia of the load is minimized by having a large reduction ratio (at least 10:1) in the first pair of gears in the speed reducer, since

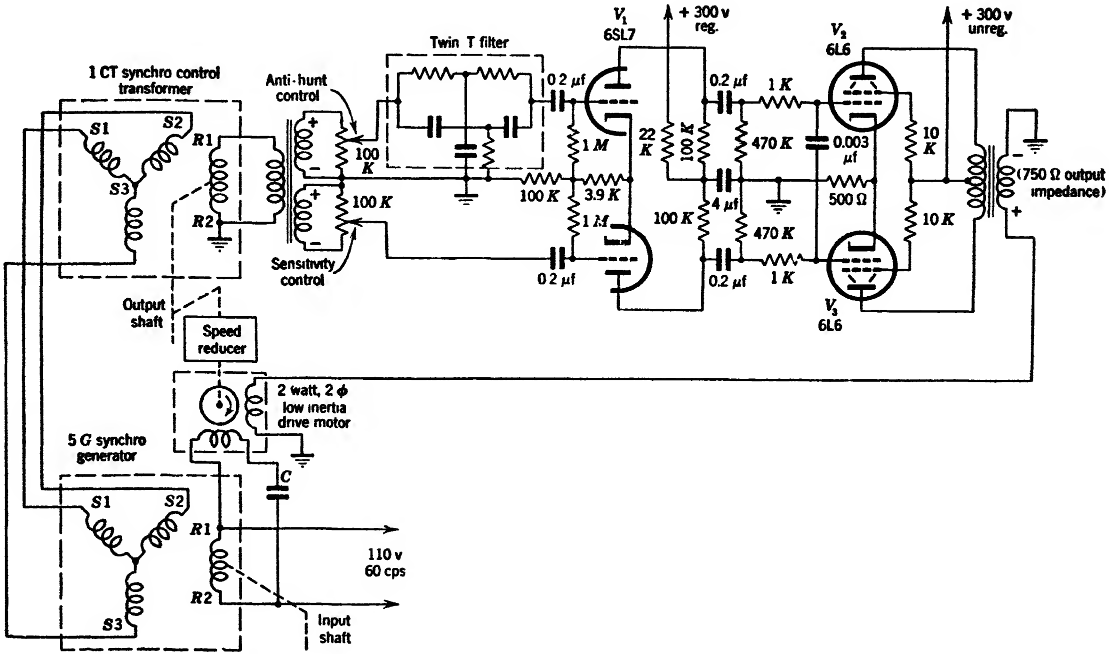


FIG. 9·7 Circuit of a low power servomechanism suitable for handling, with high precision, the light loads characteristic of many laboratory applications.

the inertial load presented by the rest of the reducer and the rotor of the CT is reduced by the square of this ratio. All sources of backlash in the mechanical system should be reduced as much as possible.

The circuit includes a frequently used scheme, involving a twin-tee network (page 236), for improving the stability of the overall system by supplying electrical damping. We have seen that the error signal goes to zero when the angular error becomes zero. Thus although no signal is applied to the drive motor at the instant it reaches its correct position, yet its momentum together with that of its load will carry it beyond the correct position, resulting in frictionally damped oscillation or hunting about the position of zero error. Obviously the inertia of the motor and its load should be made as small as possible. If the gain of the amplifier is increased in an attempt to increase the tightness of the servo system, a point will be reached where sustained oscillations occur. It can be seen that damping can be achieved by feeding to the drive motor, in addition to the error signal, the first derivative of the error signal; thus if the amplitude of the error signal is positive [5] and decreasing, its derivative will be negative and, if amplified and applied to the drive motor, will exert a restraining torque even when the error signal itself vanishes.

The function of the twin-tee filter in damping the hunting can be qualitatively understood as follows. The twin-tee is tuned to the carrier frequency of the synchro circuit, so that an error signal of constant amplitude will give no output. If, however, the amplitude of the error signal varies, so that frequencies other than f_0 are involved, the output will not vanish. In particular, if the error signal is of the form

$$e = E \sin \delta t \sin \omega_0 t = \frac{E}{2} [\cos (\omega_0 - \delta)t - \cos (\omega_0 + \delta)t]$$

the outputs resulting from the two side frequencies can be seen by reference to Fig. 7·8 (*b*) to be approximately

$$\frac{E}{2} k \, \delta \cos \left[(\omega_0 - \delta)t + \frac{\pi}{2} \right] \quad \text{and} \quad -\frac{E}{2} k \, \delta \cos \left[(\omega_0 + \delta)t - \frac{\pi}{2} \right]$$

if $\delta \ll \omega_0$, so that the two amplitudes can be set equal and the phase shift is $\pm\pi/2$. The constant k is the slope of the amplitude

[5] By a positive amplitude for the error signal is meant a zero phase difference between the error signal and the reference (synchro excitation) voltage.

characteristic for ω slightly larger than ω_0. These two outputs add to give $-Ek\,\delta \cos \delta t \sin \omega_0 t$, so that the output amplitude is proportional to the derivative of the input amplitude. A similar result is, of course, obtained if the input amplitude varies as $\cos \delta t$. Since in general $E(t)$ can be expressed as a Fourier series, the twin-tee in effect differentiates the amplitude of any error signal. Thus the twin-tee enables one to use larger direct gain in the servo-amplifier without getting into the region of instability of the over-all system.

The servomechanism of Fig. 9·7 can develop sufficient torque for the majority of shaft control problems which may arise in laboratory work, since in nearly all cases the speed of the shaft to be controlled will be very much smaller than that of the motor shaft.

A Two-Speed Servomechanism

The system of Fig. 9·7 is a 1-speed system; that is, the synchro shafts turn at the same rate as the input and output shafts. The

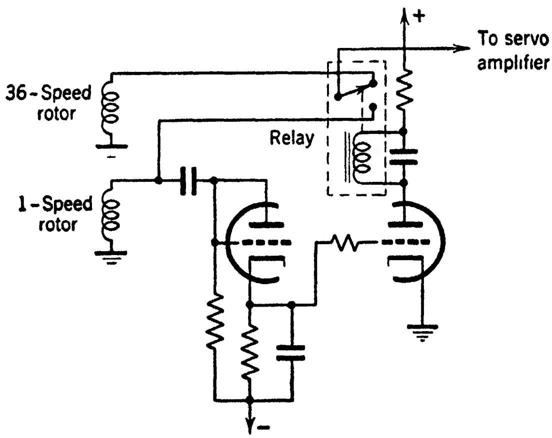


FIG. 9·8 Diagram illustrating a relay circuit for switching the amplifier of a double-speed servomechanism from the high speed to the low speed error signal whenever the latter exceeds a certain amplitude.

angular error between the input and output shafts can be reduced to as little as a few hundredths of a degree by incorporating suitable gearing to make the synchro shafts execute, say, 36 revolutions

to each revolution of the input and output shafts. Such a 36-speed system can, of course, lock in any one of 36 stable relative positions, so that further refinements are required to remove this uncertainty. One satisfactory scheme is to parallel the 36-speed pair of synchros by a 1-speed pair, and to provide means for automatically switching the servoamplifier from the 36-speed *CT* to the 1-speed *CT* whenever the amplitude of the 1-speed error signal becomes too large in either the positive or the negative direction. The switching can be done by the relay circuit shown in Fig. 9·8, as well as by other methods. The relay normally connects the 36-speed *CT* to the amplifier, but when the 1-speed error signal exceeds a certain amplitude the rectifier-integrator circuit increases the triode plate current sufficiently to actuate the relay, and the 1-speed *CT* takes control until the system has been returned to a point where the 36-speed *CT* can safely resume control.

High Power Servomechanisms

It can be seen that a servomechanism designed to operate any large load requires a large amount of power amplification. Power amplifiers, such as *amplidynes*, which operate considerably more efficiently than do vacuum tubes, are available for this purpose.

The principles involved in the design and operation of high power servomechanisms are the same as those pertaining to low power systems. However, it is much more important that the design of a high power system be sound, since what may be only an annoying hunting in a low power system may result in serious damage to equipment or personnel in a high power system.

D-C Servomechanisms

The control of a shaft position can also be obtained by a servo system in which a d-c error signal is used. In the system represented by the diagram of Fig. 9·9, voltages E_1 and E_2, defining respectively the input and output shaft positions, are compared in a differential amplifier (see, for example, page 300), and the signal proportional to the difference $E_1 - E_2$ is amplified to drive a reversible d-c motor.

A generally more satisfactory system is represented in Fig. 9·10 (*a*). Here any difference between E_1 and E_2 is converted to

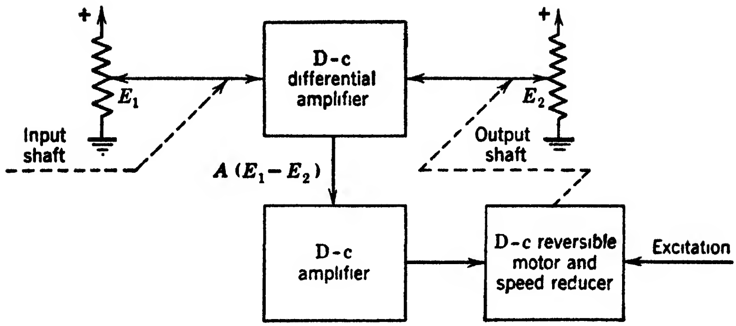
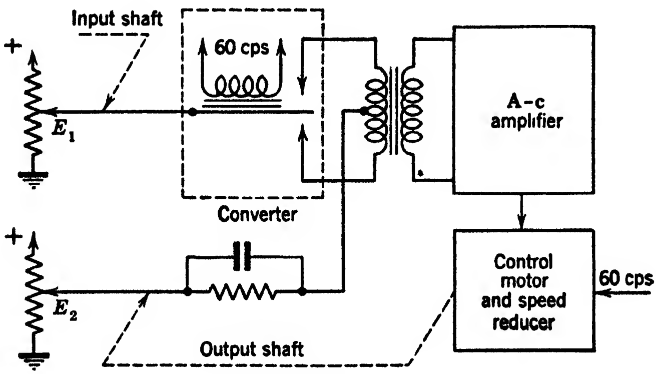
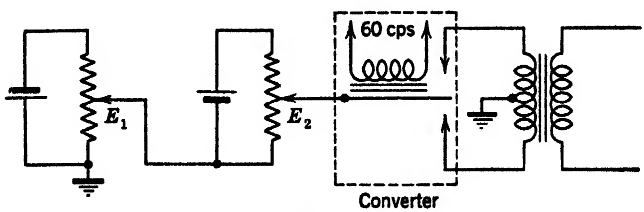


FIG. 9·9 Representation of a d-c shaft-control servomechanism. The motor is arranged to turn the output shaft so that $E_1 - E_2$ is made small.



(a)



(b)

FIG. 9·10 (a) Diagram of a d-c servo with an a-c amplifier. The use of a converter to change the d-c error signal to an a-c signal avoids the difficulties inherent in d-c amplification. (b) An alternative arrangement of the converter in the above servo.

a 60-cycle signal (actually a square wave) by a synchronous vibrating reed,[6] and the resulting signal is amplified and used to energize a two-phase a-c motor. Damping is supplied, if necessary, by the *RC* combination in the input circuit; this combination serves to differentiate the error voltage. This system avoids the necessity of d-c amplification, which is apt to be troublesome. An alternative arrangement of the converter circuit is shown in Fig. 9·14 (*b*), which may be applied when E_2 does not have a grounded terminal. It is seen that either of these arrangements has the important property that at balance no current is drawn from either voltage source by the converter. This fact has been applied in the construction of continuously self-balancing recording potentiometers [6] suitable for the measurement of very small direct voltages. In this application, E_1 is derived from a carefully calibrated slide-wire carrying a known current, and E_2 is the voltage to be recorded. The recording is done by a pen fastened to the carriage, controlled by the motor, which moves the slide-wire contactor.

Rate Servos

Servomechanisms for controlling the *rate* of rotation of a shaft by means of a direct voltage can be based on the systems represented in Figs. 9·9 and 9·10, if the potentiometer producing E_2 is replaced by a tachometer generator.

9·4 COMPUTERS

In this section we will discuss a few simple examples of computers of the electrical analogue type. The reader should be able, on the basis of the material presented here, to understand the functioning of more complex computers and to devise systems for handling some calculations not explicitly considered in the following paragraphs.

It is obvious that a single variable parameter can be represented by a direct or alternating voltage or current, or by a resistance or other electrical quantity, and that in principle the representation can be carried to any desired degree of precision. In practice it is

[6] Such a device is manufactured by the Brown Instrument Company, Philadelphia, Pa., Leeds and Northrup Company, Philadelphia, Pa., and other concerns.

difficult to achieve a precision exceeding about 0.1 per cent. If the parameter can take on positive or negative values, its sign can be represented by the sign of a direct voltage or current, or by the phase of an alternating voltage or current. In the case of alternating voltages or currents, it is evident that the use of phase angles ranging from 0° to 360° allows one to represent by a single electrical analogue any point in a plane.

Potentiometers are widely employed for obtaining variable voltages for use in computers. In some applications the degree of linearity of potentiometers is important. Potentiometers are available [7] which are guaranteed to be linear with tolerances as small as 0.025 per cent. Such potentiometers are composed of a long coil of resistance wire formed into a helix of several turns, and therefore have available shaft rotations up to as much as 15,000°, compared with about 320° available with ordinary potentiometers. It is thus possible to obtain any desired setting with much greater precision with this type of potentiometer. Alternating voltages of variable amplitude can also be obtained from autotransformers.

Addition and Subtraction

The direct procedure of connecting two voltages in series or in series opposition to evaluate their sum or difference needs no further elaboration here. For some applications, as for example when the loading imposed on the voltage sources must be kept to a minimum, electronic methods are required. The circuit [8] of Fig. 9·11 gives a d-c output voltage which is proportional to the difference between two d-c input voltages, and is unaffected by a change in the level of the two inputs provided the difference between them remains constant. Analysis of the circuit gives the following results (all voltages measured relative to ground):

$$E - \frac{E_{bb}}{2} = \frac{\mu}{\mu + 2 + \dfrac{r_p}{R_1}} \frac{E_1 - E_2}{2} = \frac{E_{bb}}{2} - E' \tag{9·1}$$

$$E - E' = \frac{\mu}{\mu + 2 + \dfrac{r_p}{R_1}} (E_1 - E_2) \tag{9·2}$$

[7] Manufactured by the Helipot Corporation, South Pasadena, Calif.

[8] J. W. Gray and D. MacRae, Radiation Laboratory Report 457, Nov. 1943.

In a similar manner, addition or subtraction of alternating or direct voltages can be accomplished by the cathode-coupled amplifier of Fig. 10·6.

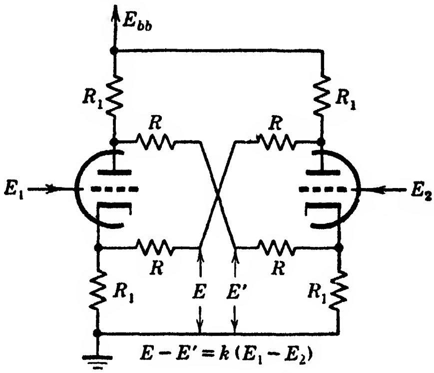


FIG. 9·11 Circuit giving the difference between two direct voltages.

Multiplication

A voltage proportional to the product of two quantities x and y results from the arrangement shown in Fig. 9·12. The first potentiometer is given a setting proportional to x and the second a setting proportional to y; the output voltage, provided $R_2 \gg R_1$, is then proportional to xy. Obviously a similar arrangement involving autotransformers can be applied to alternating voltages.

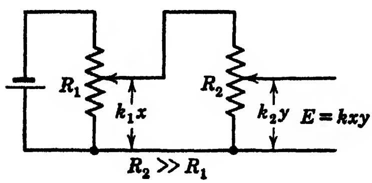


FIG. 9·12 Two potentiometers arranged for evaluating the product of two quantities.

Evaluation of Various Functions

A voltage proportional to the sine (or cosine) of an angle can be obtained with considerable precision from a so-called *sine card*. The usual form of sine card (Fig. 9·13) is a winding of fine wire on a flat card, with a contacting arm which rotates about the center of the card. It is important that many turns be employed so that the resistance of one turn is negligible compared to the total re-

sistance of the card. Sine cards are frequently used to determine the altitude of a radar target; if E_0 is proportional to the slant range as determined by the radar and θ is the elevation angle of the radar antenna when it is properly trained on the target, the output of the sine card is proportional to the altitude of the target.

Trigonometric functions may also be evaluated by means of rotary transformers. In Section 9·1 we described synchro generators and control transformers which can obviously be used for this purpose. Another important type of rotary transformer is called a *resolver*, which in its usual form consists of two rotor and two stator windings, each pair of windings being in space quadrature. If in-phase alternating voltages of amplitudes E_{R1} and E_{R2} are applied to the rotor windings, the amplitudes of the stator voltages are given (with proper definition of signs) by the equations

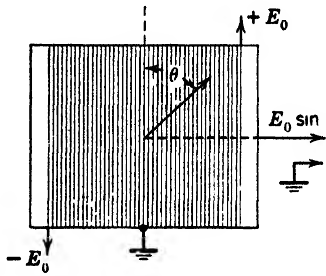


FIG. 9·13 A sine card for evaluating sin θ or cos θ.

$$
\begin{aligned}
E_{S1} &= E_{R1}\cos\theta + E_{R2}\sin\theta \\
E_{S2} &= -E_{R1}\sin\theta + E_{R2}\cos\theta
\end{aligned}
\qquad (9\cdot 3)
$$

As with synchros, these output voltages are of course in phase with the exciting voltages.

If variable voltages, for example proportional to range, are fed to a resolver, such voltages should not be too small since resolvers are non-linear at low voltages. Care must be taken not to impose a low impedance load on a resolver if accurate results are desired. Obviously, if in equations 9·3 one sets $E_{R1} = 0$ and $E_{R2} = kr$, one can convert r, θ to x, y. If one resolver is used to drive another, with a linear buffer amplifier between them, one can convert r, θ, ϕ to x, y, z.

Appropriately loaded potentiometers may be used for approximating a variety of functions. For example, the circuit [9] shown

[9] Cf. R. Hofstadter, *Rev. Sci. Instruments*, **17**, 298 (1946). See also L. A. Nettleton and F. E. Dole, *ibid.*, **18**, 332 (1947).

in Fig. 9·14 gives an approximation to the tangent function having an error of less than ±0.5 per cent for angles between −1 and +1 radian. It is easily found that the output voltage is

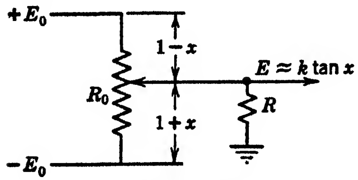


FIG. 9·14 A potentiometer loaded for approximating the sine function.

$$E = \frac{xE_0}{1 + \frac{R_0}{4R}(1 - x^2)} \tag{9·4}$$

where the quantities have the significance shown in the figure. If $R_0/4R$ is set equal to $\alpha/(1 - \alpha)$, we have

$$\frac{E}{E_0(1 - \alpha)} = x + \alpha x^3 + \alpha^2 x^5 + \cdots \tag{9·5}$$

Comparison of this expression with the series expansion of $\tan x$,

$$\tan x = x + \frac{1}{3}x^3 + \frac{2}{15}x^5 + \cdots \tag{9·6}$$

suggests that α should be set equal to ⅓. Actually a somewhat better approximation is obtained if $\alpha = 1/2.792$.

Computers Employing Servomechanisms

The scheme for multiplication illustrated in Fig. 9·12 requires the setting of at least one potentiometer to represent an input quantity. In applications where it is necessary to have this done automatically a servomechanism is employed. The use of a servomechanism also permits carrying out the inverse operation, in this case division.

Consider the general scheme outlined in Fig. 9·15. Direct input voltages E_1, E_2, and E_3 are used, one or more of which may be fixed voltages. The potentiometers, with ganged shafts, may have linear or non-linear tapers, or may be loaded as in Fig. 9·14. The servomechanism is a zeroing servo of the type illustrated in Fig. 9·9 or 9·10, so that it adjusts the potentiometer shaft position to make the output of potentiometer P_1 equal to E_3. Let E'_1

$= E_3 = E_1 f_1(\theta)$ and $E_{out} = E_2 f_2(\theta)$. A variety of computations can be performed by this system, some of which are outlined in

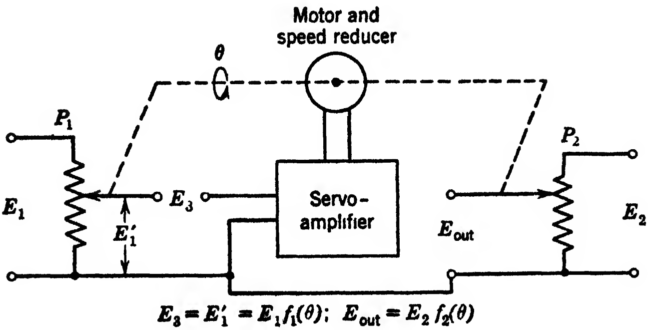


FIG. 9·15 Schematic representation of a computer using a servomechanism for adjusting potentiometer P_1 to make $E'_1 = E_3$.

Table 9·1. The reader can easily see the limitations which must be imposed on the values taken by the various input voltages.

TABLE 9·1 ILLUSTRATING THE COMPUTATIONS PERFORMED BY THE SYSTEM IN FIG. 9·15

E_1	E_2	E_3	$f_1(\theta)$	$f_2(\theta)$	E_{out}	Notes
1	E_a	E_b	$f_1(\theta)$	$f_1(\theta)$	E_aE_b	(a)
E_a	1	E_b	$f_1(\theta)$	$f_1(\theta)$	E_b/E_a	(a)
E_a	E_b	1	$f_1(\theta)$	$f_1(\theta)$	E_b/E_a	(a)
E_{out}	1	E_a	$f_1(\theta)$	$f_1(\theta)$	$\sqrt{E_a}$	(a) (b)
E_a	1	E_b	$\sin\theta$	$k\theta$	$k \sin^{-1}(E_b/E_a)$	(c)
1	1	E_a	$k\theta$	$\sin\theta$	$\sin(E_a/k)$	(c)
$1/(1-\alpha)$	1	E_a	$\tan\theta$	$k\theta$	$k \tan^{-1} E_a$	(d)

(a) In cases where $f_1(\theta) = f_2(\theta)$, it will usually be most convenient to use linear potentiometers.

(b) The resistance of P_1 must be large compared to that of P_2.

(c) A sine card (Fig. 9·19) is used.

(d) $\alpha = 1/2.792$; P_1 is a loaded potentiometer (Fig. 9·20).

The computer illustrated in Fig. 9·16 illustrates another application of a zeroing servomechanism in performing the inverse of an operation which can itself be accomplished directly. We saw on page 289 how a resolver can be employed to convert polar coordinates r, θ to Cartesian coordinates x, y. If voltages propor-

tional to x and y are fed to the stator windings and the rotor is turned until $E_{R1} = 0$, E_{R2} will be proportional to r and the shaft position will give θ. In the computer of Fig. 9·16 the rotor is

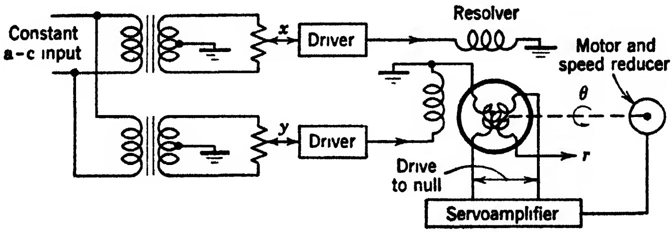


Fig. 9·16 A system for automatically converting x, y to r, θ. If the input shafts are set to give voltages proportional to x, y, the voltage across one of the resolver rotor windings is proportional to r, and the position of the rotor shaft equals θ.

turned by a servo system which may be similar to those described earlier in this chapter.

Figure 9·17 shows a scheme for solving a right triangle, which is due to R. M. Walker of the Radiation Labora-

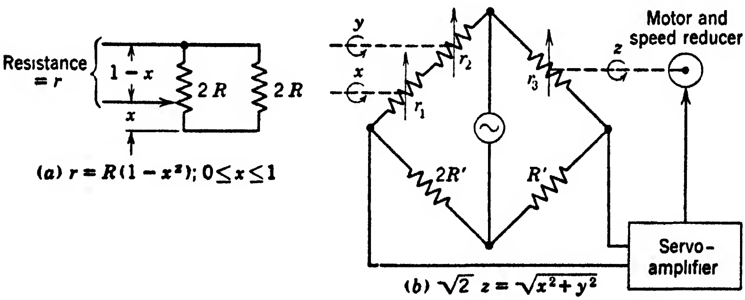


Fig. 9·17 Diagram of a right-triangle solver. The variable arms of the Wheatstone bridge are composed of loaded potentiometers of the type shown in (a). When the bridge is balanced by the servomechanism, z is proportional to the hypothenuse of a right triangle of sides x and y.

tory. The network of Fig. 9·17 (a) is a variable resistance having the value $r = R(1 - x^2)$, if $0 \leq x \leq 1$. Three such networks, $r_1 = R(1 - x^2)$, $r_2 = R(1 - y^2)$, $r_3 = R(1 - z^2)$, are connected in an a-c Wheatstone bridge as shown in Fig. 9.17 (b). The condition for balance is $z = (1/\sqrt{2})\sqrt{x^2 + y^2}$, a servo

system being used for balancing the bridge by rotation of the z shaft.

Integration of a varying voltage can be accomplished by means of a rate servo (page 286), provided the overall linearity of the servo is sufficiently good. It is merely necessary to count output shaft revolutions during the integration period.

PROBLEMS

9·1 Two independent shafts rotate at 15 rpm approximately in synchronism, so that their relative position is never more than $\frac{1}{10}$ revolution from its proper value. Give a detailed block diagram of a servomechanism for continuously indicating their relative position with an accuracy of $\pm 0.1^{\circ}$.

9·2 Derive equation 9·4.

9·3 Show how the following quantities can be evaluated by the system shown in Fig. 9·16, or a modification thereof:

(a) $\dfrac{E_a E_b}{E_c}$ (b) $\sqrt{E_a E_b}$ (c) E_a^3

CHAPTER 10

MISCELLANEOUS CIRCUITS

In some of the preceding chapters we have had occasion to discuss numerous types of circuits which are important in radar practice. In the present chapter we shall take up several additional important and useful circuits, some of which were well-known before the war. It is obvious that we will have to make a more or less arbitrary selection of circuits to be considered, since a tremendous variety are of importance to radar.[1]

As has been done elsewhere in this book, specific circuit constants are given in some cases, since it is believed that such detailed information will be useful to those who may wish to employ the circuits but are not familiar with the principles of vacuum tube circuit design. In most of these cases the particular constants given are not at all unique.

10·1 REGULATED POWER SUPPLIES

For many purposes it is important to have supply voltages which are relatively unaffected by changes in line voltage or load. Thus, high precision in a range circuit (Chapter 5) requires among

[1] A very thorough treatment of radar circuits will be found in the *Radiation Laboratory Technical Series*. More complete discussions of some of the circuits mentioned in this chapter can be found in standard reference works, such as: F. E. Terman, *Radio Engineers' Handbook*, McGraw-Hill Book Co., 1943; H. J. Reich, *Theory and Applications of Electron Tubes*, McGraw-Hill Book Co., 1939; V. K. Zworykin and G. A. Morton, *Television*,

other things careful regulation of the power supply for the circuit. Filament and other a-c supplies may be stabilized by means of regulating transformers,[2] provided the line frequency is sufficiently constant.

Regulation by Gas Discharge Tubes

The voltage drop in a gas discharge is relatively independent of the discharge current over considerable ranges. This fact is utilized in voltage-regulating tubes such as the VR-75, 90, 105,

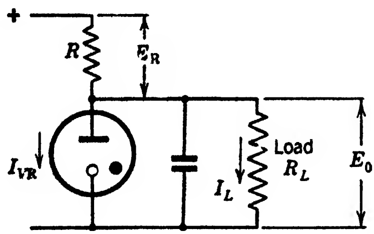


FIG. 10·1 Circuit employing a cold cathode discharge tube for obtaining a constant voltage.

and 150. These are cold cathode tubes having voltage drops, approximately given by the numbers in their designations, which change less than 5 per cent when the discharge current changes from 5 to 40 milliamperes. These tubes can thus serve as current reservoirs. The VR tube is connected to a higher voltage source through an appropriate dropping resistor R as shown in Fig. 10·1, and is usually bypassed by a condenser. As a rule the operating point will be chosen in approximately the center of the VR tube characteristic, with the VR current about 20 milliamperes, so that both decreases and increases of supply voltage or load resistance can be accommodated. Output voltages higher than 150 volts

John Wiley and Sons, Inc., 1940; O. S. Puckle, *Time Bases*, John Wiley and Sons, Inc., 1943; E. E. Zepler, *The Technique of Radio Design*, John Wiley and Sons, Inc., 1943; J. G. Brainerd, G. Koehler, H. J. Reich, and L. F. Woodruff, *Ultra-High-Frequency Techniques*, D. Van Nostrand, 1942.

[2] It should be noted that the output of commonly available regulating transformers is a distorted sine wave, which may be objectionable in some applications.

can be obtained by connecting tubes in series. VR tubes cannot be connected in parallel because of the variations from tube to tube in striking voltage.

Electronic Regulation

Degenerative feedback circuits are frequently employed in power supplies to give very low effective internal impedance and very slight dependence on supply voltage. A typical circuit is shown in Fig. 10·2. A constant reference voltage is supplied by

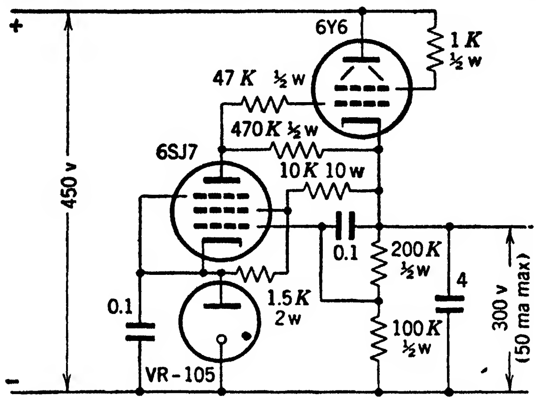


Fig. 10·2 Diagram of a typical voltage regulating circuit. The 6SJ7 amplifies changes in the output voltage and applies the amplified changes degeneratively to the 6Y6 control tube.

the VR-105. If for any reason the output voltage tends to rise, the grid of the 6SJ7 amplifier will rise and its plate will fall. This will cause the drop in the 6Y6 control tube to increase, and thus to oppose the change in the output voltage. The 6SJ7 screen grid is operated at an unusually low potential so that the plate current may have the requisite low value at grid voltages which are well away from the nearly horizontal portion of the grid voltage-plate current characteristic. The condenser across the output is used to smooth out rapid fluctuations which cannot be degenerated by the feedback loop because of the large amplifier load resistor. If higher current output is required, several control tubes can be connected in parallel; if this is done small resistors should be placed in series with the grid, screen, and plate of each tube to damp out high frequency oscillations.

It should be emphasized that electronic regulation does not add as much to the cost of a power supply as might at first be supposed

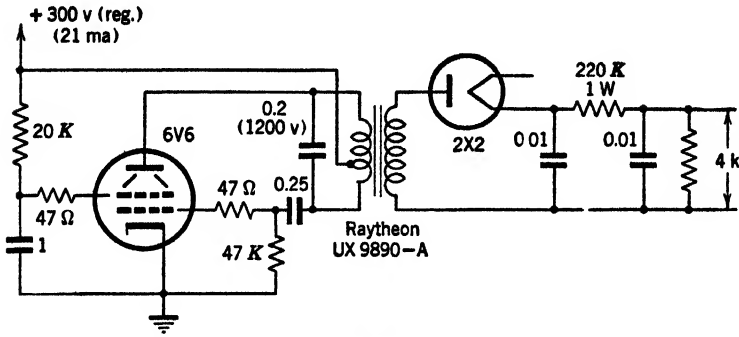


FIG. 10·3 A low current, high voltage regulated power supply suitable for light loads such as cathode ray tube anodes.

since less filtering of the high voltage fed to the regulator is required, a single-section LC filter being sufficient in most cases.

A supply such as that shown in Fig. 10·2 can obviously be used to give a regulated negative voltage if its positive output terminal is grounded.

For some purposes it is desirable to regulate the intensification voltage applied to a cathode ray tube (Chapter 6). In this application it is generally permissible to use a power supply with a rather high internal impedance, since the load currents are very small. An ingenious method for obtaining protection of the high voltage against fluctuations in the a-c supply voltage is illustrated by the circuit of Fig. 10·3. This circuit, which was developed at the Radiation Laboratory, consists of an approximately 200-cycle oscillator which is powered by a well-regulated 300-volt supply. The tank circuit of the oscillator is the primary of a transformer

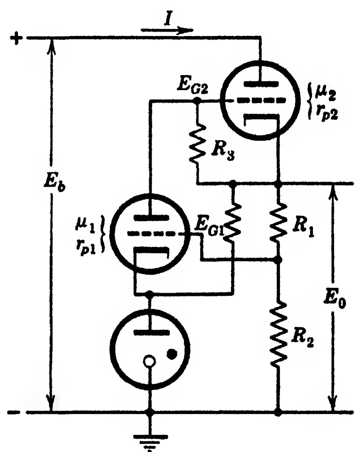


FIG. 10·4 Simplified schematic diagram of a degenerative voltage regulating circuit.

from the secondary of which, after suitable rectification and filtering, the high voltage is derived.

The performance of a regulated power supply may be expressed by its internal resistance R_0 and its gain A_0. If E_b is the input voltage (Fig. 10·4), E_0 the output voltage, and I the output current (neglecting the current through R_1, R_2, and the VR tube) R_0 and A_0 are defined by the expressions

$$A_0 = \frac{dE_0}{dE_b} \text{ (constant load)}$$

$$R_0 = \frac{dE_0}{dI}$$

It is desired to have both of these derivatives small. It can be shown that the following equations are approximately true if $\mu_2 A_0 \ll 1$:

$$A_0 = \frac{r_{p1} + R_3}{\mu_1 \mu_2 R_3} \frac{R_1 + R_2}{R_2} \tag{10·1}$$

$$R_0 = A_0(r_{p2} + R) \tag{10·2}$$

where R is the internal impedance of the unregulated supply from which E_b is taken, and the remaining quantities have the significance shown in Fig. 10·4. For the circuit of Fig. 10·2, $A_0 \approx 0.005$ and $R_0 \approx 7.5$ ohms, if $R \approx 500$ ohms.

More elaborate regulating circuits have been devised which give closer regulation than the one shown in Fig. 10·2, though the latter is satisfactory for most purposes. Lawson [3] has described a power supply designed to deliver up to 350 milliamperes at 250 volts which has an internal impedance of only 0.05 ohm. The output voltage of this supply changes less than 5 millivolts for an a-c input change from 105 to 125 volts.

10·2 CATHODE FOLLOWER CIRCUITS

The load which a vacuum tube amplifier or other circuit is required to drive is frequently of low impedance, which requires that the output impedance of the circuit should also be low. In low frequency circuits such a load is frequently coupled to its

[3] J. L. Lawson, Radiation Laboratory Report 44, Feb. 26, 1945.

driver by an impedance-matching transformer. Since transformers are not as yet available of sufficiently wide pass band to serve in video circuits, low output impedances must be obtained by other means. The *cathode follower* circuit, illustrated in Fig. 10·5 has been widely used for this purpose. The output impedance R_0 of such a circuit can be shown to be approximately equal to the resistance of a parallel combination of the cathode resistor R_K and a resistor equal to the reciprocal of the transconductance of the tube:

$$R_0 = \frac{R_K}{1 + g_m R_K} \tag{10·3}$$

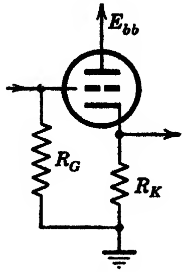


FIG. 10·5 Basic cathode follower circuit.

Thus the output impedance of a 6AC7 cathode follower is less than 100 ohms. (It should be noted that in some circuits the plate current in a cathode follower is made to vary over a large range, so that the g_m cannot be assumed to be constant; in the limit, if the plate current goes to zero, the output impedance becomes simply R_K.)

The voltage gain of a cathode follower is

$$A = \frac{\mu R_K}{r_p + (\mu + 1)R_K}$$

or, if $\mu \gg 1$, as is usually the case,

$$A = \frac{g_m R_K}{1 + g_m R_K} \tag{10·4}$$

This expression shows that the voltage gain of a cathode follower is always less than unity, approaching unity as $g_m R_K$ becomes much larger than unity. However, the *power gain* may be very large, since very little grid signal power is in general required.

The inverse feedback inherent in a cathode follower leads to good amplitude and phase characteristics up to high frequencies, and allows large input signals without overloading. The feedback also reduces the effective input capacity to the value $C(1 - A)$, where C is the actual input capacity. The input resistance is R_G, unless the grid resistor is returned to the cathode instead of ground,

in which case the input resistance is approximately $(\mu + 1)R_G$ (if $g_mR_K \gg 1$). These facts concerning the input impedance of a cathode follower have the important result that it is possible to parallel several such tubes to increase the power output without imposing a low impedance load on the driver stage preceding the output tubes.

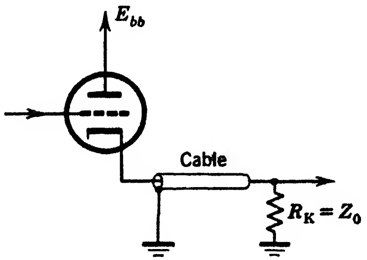


FIG. 10·6 Illustrating the use of a cathode follower to drive a low impedance coaxial line.

Figure 10·6 shows a common method of using a cathode follower to drive a low impedance coaxial cable. Proper termination of the far end of the cable ensures the absence of reflections in the cable, and the terminating resistor serves as the cathode resistor for the tube. There is in general an impedance mismatch at the tube end of the cable, which can be removed by an appropriate series or shunt resistor; this mismatch usually is not objectionable since

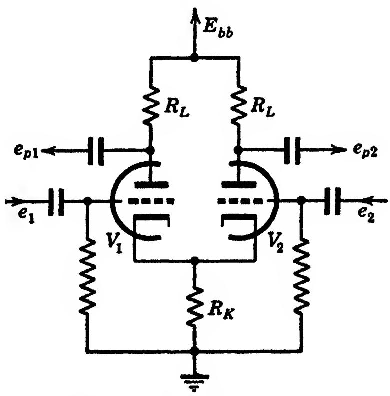


FIG 10·7 Schematic diagram of a cathode-coupled amplifier, useful for phase splitting and mixing operations.

there is no chance for troublesome reflections to occur here if none occurs at the far end of the cable.

An interesting application of the cathode follower is the cathode-coupled amplifier ("long-tailed pair") shown schematically in

Fig. 10·7. The common cathode resistor is usually taken large enough so that $(\mu + 1)R_K \gg r_p + R_L$. In this case, if V_1 and V_2 have the same characteristics, the output voltages are as given in Table 10·1. A_0 is the voltage gain to be expected for a single triode with no cathode degeneration:

$$A_0 = \frac{\mu R_L}{r_p + R_L} \tag{10·5}$$

TABLE 10·1 OUTPUT SIGNALS OBTAINED FROM THE CATHODE-COUPLED AMPLIFIER OF FIG. 10·7

Case	Input Signals	Type of Operation	Output Signals
a	$e_1 = e_2$	In-phase input signals	$e_{p1} = e_{p2} = 0$
b	$e_1 = -e_2$	Push-pull amplifier	$e_{p1} = -e_{p2} = -A_0e_1$
c	$e_2 = 0$	Phase splitter	$e_{p1} = -e_{p2} = -\frac{1}{2}A_0e_1$
d	$e_1 \neq e_2$	Mixer and phase splitter	$e_{p1} = -e_{p2} = -\frac{1}{2}A_0(e_1 - e_2)$

Case a shows that in-phase signals, such as hum signals at the cathodes, are degenerated. Case b is useful for rebalancing push-pull signals. Case c gives a very satisfactory method for splitting a "single-ended" signal into push-pull signals without using a transformer. Case d combines mixing two independent signals and phase splitting to give balanced push-pull outputs.

10·3 MULTIVIBRATORS

Multivibrators are widely used as frequency dividers or multipliers (hence the name), gate generators, and delay circuits. Numerous more or less distinct types are known; we shall describe three simple types which are useful in radar applications.

Basic Multivibrator

The basic circuit of a multivibrator is shown in Fig. 10·8. It is composed of two regeneratively intercoupled amplifier stages. Owing to the regeneration the circuit is only stable with one tube conducting and the other one cut off. Suppose V_1 is conducting and V_2 is cut off. The charge on condenser C_1 will gradually leak off, allowing the grid of V_2 to come up to the cutoff value. As soon as V_2 starts to conduct, the potential on its plate decreases, and because of the capacitive coupling the potential of the grid of V_1 also decreases. This causes the potential of the plate of V_1 to

rise, and this rise is passed on to the grid of V_2 by C_1. Thus the grid of V_2 rapidly rises to zero potential, where grid current prevents further rise, and the grid of V_1 is driven far beyond cutoff. The potential of the grid of V_1 now gradually rises toward cutoff at a rate determined largely by the time constant R_1C_2; when it reaches cutoff the situation is rapidly reversed and the plate current shifts from V_2 back to V_1.

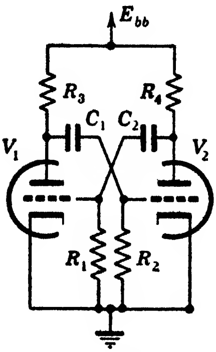


FIG. 10·8 Basic multivibrator circuit. The two amplifiers are regeneratively intercoupled, with the result that at any instant one tube will be conducting and the other cut off, with periodic reversals of this situation.

The waveforms associated with this circuit are shown in Fig. 10·9. If the two halves of the circuit are identical, the on periods are equal; they can be made unequal by suitable selection of components. The drop of plate potential is very rapid, requiring only a microsecond or less, whereas the rise is relatively slow because of the time constant composed of the load resistor and various capacities, chiefly the coupling capacitor C_1 which was largely discharged while the plate potential was low. On the other hand, the grid fall and the final grid rise are both very rapid (less than a microsecond). The plate rise can be made faster by using video techniques as described in Chapter 5 (i.e., decreasing the load resistors and using tubes with low input capacities and high current-carrying ability). For example, if the output is to be taken from the plate of V_2, one might employ a 6V6 for V_2 so that a large output would be available with a relatively small load resistor, and a 6AG7 for V_1. The grid rise and fall between zero volts and cutoff, which is very rapid, is large enough for many purposes; for example, this is a very convenient range of voltage for gating purposes (see Section 10·6) when the gating pulse is applied to the grid, or cathode (through a cathode follower), of a sharp cutoff tube such as a 6AC7 or 6SL7. In such a case direct coupling from the multivibrator grid may be used.

The multivibrator in Fig. 10·8 is free-running. However, it can be synchronized with a repetitive waveform, such as a sine wave, of frequency somewhat higher than the natural frequency of the

multivibrator, or a multiple thereof. It is evident that if a small positive voltage is applied to the grid of V_2 just before it reaches cutoff, the multivibrator will undergo its shift of plate current. Such a synchronizing voltage can be applied to both grids. Since the output of a multivibrator is rich in harmonics, the circuit can be used for frequency multiplication as well as for frequency division.

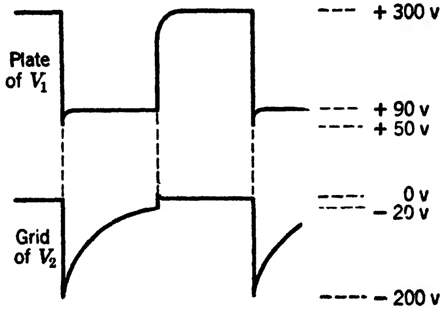


FIG. 10·9 Approximate waveforms for the multivibrator of Fig. 10·8, obtained by using the following circuit constants:

$E_{bb} = 300$ v $\qquad C_1 = C_2 = 0.001\ \mu f$
$R_1 = R_2 = 470K \qquad V_1 = V_2 = \frac{1}{2}$6SN7
$R_3 = R_4 = 24K$

The on periods of V_1 and V_2 can be rather closely approximated by the expressions

$$t_1 = C_1R_2 \ln \frac{E_{bb} - E_1}{E_{C2}} \tag{10·6}$$

$$t_2 = C_2R_1 \ln \frac{E_{bb} - E_2}{E_{C1}} \tag{10·7}$$

where t_1, t_2 are the on times of V_1 and V_2 respectively; E_1, E_2 are the plate potentials of V_1 and V_2 with the coupling condensers disconnected; E_{C1}, E_{C2} are the cutoff potentials of V_1, V_2; and E_{bb} is the supply voltage. These expressions hold if $R_2 \gg r_{p1}$ and $R_1 \gg r_{p2}$, where r_{p1}, r_{p2} are the plate resistances of V_1, V_2.

Biased Multivibrator

For some purposes it is necessary for the multivibrator to go through just one cycle of operation initiated by an external trigger, and then to remain in its original state until another trigger

comes along. This is accomplished (Fig. 10·10) by replacing one of the coupling condensers of Fig. 10·8 by a "direct" or resistive coupling, and by returning the grid to which this coupling goes to a negative voltage instead of to ground. If the components are properly chosen, the direct-coupled grid will be held well below cutoff. If a positive trigger is applied to this grid or, better, if a negative trigger is applied to the other grid, a regenerative shift of plate current takes place, which is again reversed as in a stand-

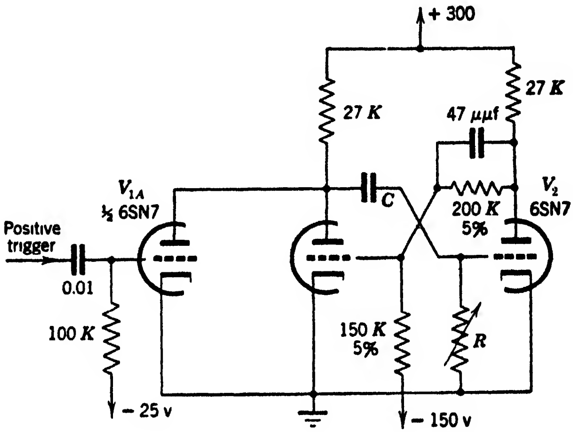


FIG. 10·10 Multivibrator circuit with one tube biased below cutoff, so that no change takes place until an external trigger is applied.

ard multivibrator when the capacitively coupled grid reaches cut-off after an interval determined by the time constant RC (Fig. 10·10). Such a multivibrator is variously termed a biased, one-shot, or delay multivibrator. The small bypass condenser supplies a low impedance path to the grid of V_{2A} for the high frequency components of the plate changes of V_{2B}. The waveforms given in Fig. 10·9 apply qualitatively to this circuit, except that after the first plate returns to E_{bb} there is no further change until the time of the next trigger.

The length of the "firing" interval of the multivibrator of Fig. 10·10 is given approximately by equation 10·6 (with appropriate changes in symbols); this length can be conveniently changed by varying R. This property makes this circuit convenient for producing a "delayed" trigger; in this application, the output of

either plate is coupled to the grid of a following tube by means of a differentiating circuit (page 166).

For some purposes it may be desired to have the firing period of the multivibrator occupy a rather large fraction of the time between triggers. Circuits similar to that in Fig. 10·10 cannot be used for duty cycles above about 0.7 because of the slow recovery

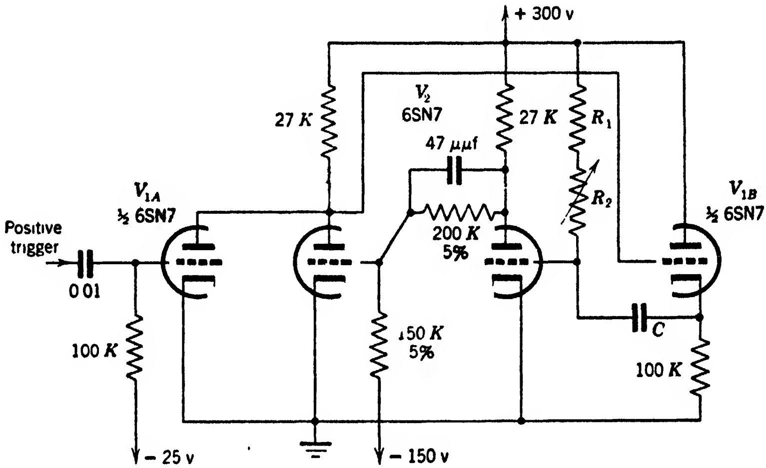


FIG. 10·11 Biased multivibrator with cathode follower recharging of the coupling condenser to permit high duty cycle operation. The grid resistor of the unbiased tube is returned to E_{bb} to decrease the variation in the time at which the grid of V_{2B} reaches cutoff.

indicated in the first curve of Fig. 10·9. The steps mentioned above which may be taken to decrease this recovery time also allow the duty cycle to be increased. The same result may be accomplished by the modification shown in Fig. 10·11. Here a rapid recharge of the coupling condenser C is brought about by using a cathode follower between the first plate and the condenser; this furnishes a low impedance path between the plate supply and the coupling condenser when the cathode follower grid potential is suddenly raised at the end of the firing period. With this arrangement the duty cycle can be increased to 0.9 or 0.95.

The circuit of Fig. 10·11 also illustrates another important point. The grid of the second tube of the multivibrator is returned to E_{bb} instead of to ground, so that as this grid starts to recover after

being driven far beyond cutoff its potential follows an exponential law which approaches asymptotically to E_{bb} rather than to ground, and it will thus strike the cutoff value at a much steeper slope. The time "jitter" of the trailing edge of the multivibrator output, which results from a-c voltages on the grid arising from heater-cathode leakage, power supply ripple, and the like, is much reduced by this procedure since these small a-c voltages change the time of reaching cutoff by amounts roughly inversely proportional to the slope of the grid potential at this time. Obviously this procedure can be employed only if $R_1 + R_2$ is large enough to limit the grid current to a safe value.

Eccles-Jordan Circuit

The Eccles-Jordan circuit shown in Fig. 10·12 is derived from a multivibrator by using direct couplings in place of both capacitive couplings. This circuit, which is also called a "flip-flop" or

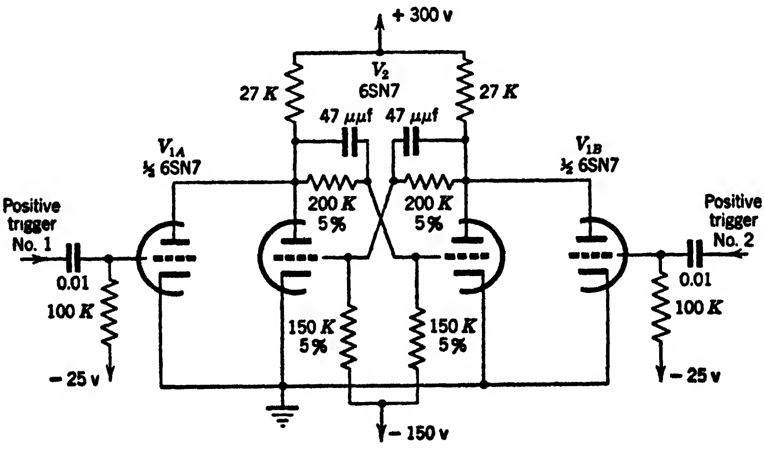


FIG. 10·12 A typical Eccles-Jordan circuit. Because both tubes are biased a complete cycle of operation requires the introduction of two external triggers.

"scale-of-two" circuit has two stable conditions. The plate current is shifted to the second tube in the multivibrator part of the circuit by applying a negative trigger to the first plate or the second grid, and is shifted back by applying a negative trigger to the second plate or the first grid. These triggers may be introduced

through buffer stages as indicated. Small bypass condensers increase the speed of the grid changes.

If positive triggers are applied to both buffer tubes simultaneously, a shift of plate current will be caused by each successive trigger. If the output at either plate is differentiated and used, for example, to "fire" a blocking oscillator, triggers will be obtained at the time of alternate input triggers. Thus the circuit serves as a scale-of-two counter.

10·4 CLAMPING CIRCUITS[4]

Clamping tubes are essentially electronic switches which connect two points for a specified period. A clamp differs from an ordinary switch in several important respects:

(a) A clamp can be opened or closed very rapidly, in a fraction of a microsecond if necessary.

(b) When closed its impedance is usually in the range 300 to 3000 ohms; when open this increases to several megohms.

(c) In some cases the clamp can only connect two points, 1 and 2, if point 1 is at a higher potential than point 2; such clamps are called *single-ended.* *Double-ended* clamps do not have this limitation.

It will not be possible to enter into any extended discussion of clamping tubes and their properties. We shall merely give a few examples of circuits where clamps are used.

D-C Restoration

Consider the video amplifier of Fig. 10·13, designed to amplify positive pulses, with the diode removed. Suppose that the signal introduced on the grid is a large positive pulse of low duty cycle (cf. page 146). Then the average voltage will be very nearly the base voltage; since the average voltage is constrained to be E_{cc}, the output from the amplifier will be as shown in Fig. 10·14 (*a*). Now if the duty cycle is increased to 50 per cent, the average voltage will become half the pulse height, and the output will be as shown in Fig. 10·14 (*b*). It is evident that the amplification de-

[4] C. W. Sherwin, Radiation Laboratory Report 572, May 1944.

creases as the duty cycle increases to 50 per cent. (The operating point is taken on the lower part of the grid-plate characteristic so that large positive pulses can be accommodated without saturation.) If now the diode is inserted, it has the effect of forcing the grid voltage to rise from the value E_{cc} on each signal regardless of the duty cycle, Fig. 10·14 (*c*).

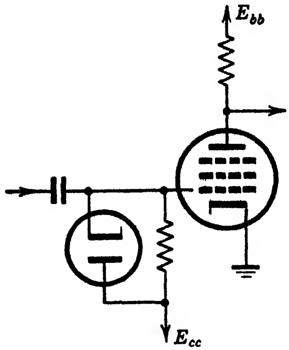


FIG. 10·13 A video amplifier stage with diode d-c restoration.

D-c restoration is also important in cases where some signals may be large enough to cause the flow of grid current, since then a negative charge is produced on the coupling condenser which can be removed rapidly only by a low impedance path such as is formed by the diode. This avoids the loss of weak signals after a large saturating one. Another use arises in the case of pulses so long that the input coupling time constant allows considerable droop dur-

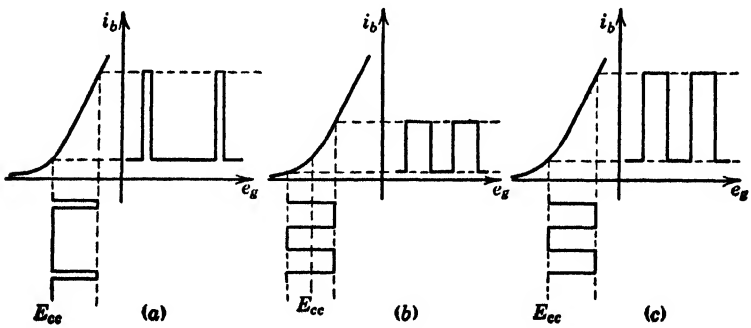


FIG. 10·14 Illustrating the use of d-c restoration to prevent a change of amplification with change in the duty cycle of the input signal. (*a*) Low duty cycle signal, with full amplification; (*b*) high duty cycle signal without d-c restoration; (*c*) the same signal with d-c restoration, full amplification regained.

ing the pulse (Fig. 10·15). In such a case the diode clamp removes the undershoot at the end of the pulse, and prevents the loss of weak signals immediately following the long pulse

It is evident that d-c restoration is of importance only when the amplitude of the input signals is an appreciable fraction of the dynamic range of the stage.

Other examples of d-c restoration have been mentioned in earlier chapters. Some of these (cf. page 215) involve the use of double-ended clamps (see below).

Recently germanium crystal rectifiers have been developed which have low "forward" impedances and large "backward" impedances for the potential differences met with in such d-c restoration applications as described above. The use of these crystals will obviously result in important savings in space and filament power as compared with clamping tubes.

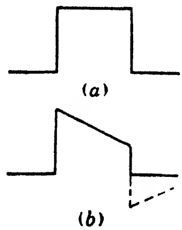


FIG. 10·15 Removal of the undershoot at the end of a long pulse by means of a diode connected as shown in Fig. 10·13. (a) Input signal; (b) signal at the grid; dotted line, without diode; solid line, with diode.

Demodulation

The circuit of Fig. 10·16 (a) illustrates the application of a double-ended clamp to the demodulation of amplitude-modulated pulses. Each signal pulse should start a little before and end a little after the corresponding clamping pulse. The demodulating action can be understood from Fig. 10·16 (b and c). The same circuit has been used for the periodic sampling of a varying audio signal in the pulse code modulation system of communication.[5]

The double-ended clamp may also be used for phase comparisons and for detecting the envelope of amplitude-modulated sinusoidal signals. The clamping signal in this case is a sine wave of the same frequency as the input signal, only the positive crests of the clamping signal serving to close the clamping tubes. The d-c output voltage changes from positive to negative as the phase of the input changes through 180°. An example of the use of this circuit for phase-sensitive detection is given on page 216.

Another circuit for detecting amplitude modulation of short pulses is shown in Fig. 10·17. This circuit employs what amounts to a single-ended clamp; the condenser in the cathode circuit of

[5] L. A. Meacham and E. Peterson, *Bell System Tech. J.*, **27**, 1 (1948).

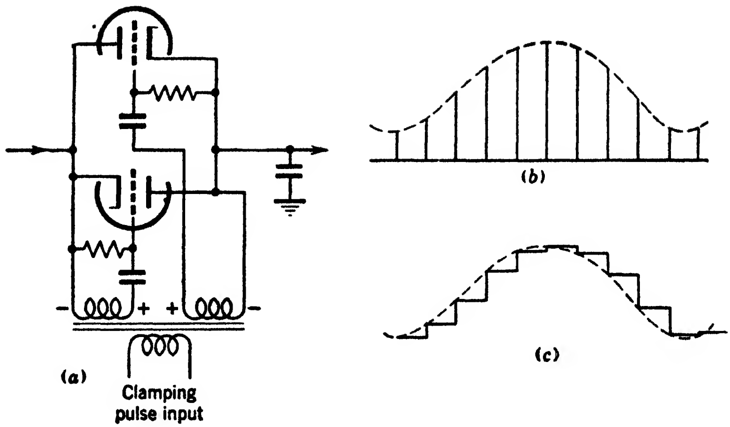


FIG. 10·16 Demodulation by means of a double-ended clamp. The output is connected to the input during each clamping pulse, so that the output condenser is charged or discharged to the pulse amplitude. The modulated pulse input is shown in (*b*). As indicated in (*c*), which represents the output, there is very little leakage from the output condenser if the output load has a high impedance.

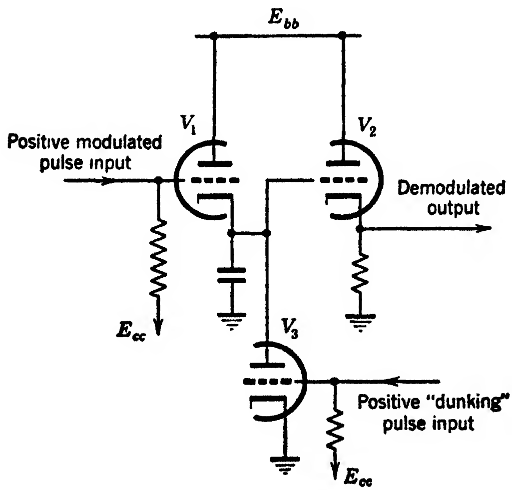


FIG. 10·17 A demodulating circuit for detecting amplitude modulation of short pulses. The limitation imposed by the fact that V_1 is a single-ended clamp is removed by "dunking" the charge on the condenser just before a new pulse arrives at the grid of V_1.

the clamping tube is cleared just before the arrival of the next pulse by means of a "dunking" pulse applied to the grid of a second clamp which discharges the condenser to ground. This is the basic circuit which was used in the work of Lawson [6] and others in studying modulation of radar echoes from various types of targets produced by various methods. In such applications, the signal to be studied is selected by means of a video amplifier which is gated on for an interval of a few microseconds at a time after the radar trigger corresponding to the range of the target. The dunking pulse precedes the gate by a few microseconds. With a carefully designed receiver and demodulation circuit it has been found possible to detect an amplitude modulation of a radar echo occurring at a single low audiofrequency amounting to as little as a few hundredths of a per cent.

10·5 BLOCKING OSCILLATOR COUNTING CIRCUITS

Numerous situations arise in which it is necessary to reduce the frequency of a pulsed signal, or actually to count the number of pulses arriving in an interval of time. A familiar example [7] is the counting of the Geiger counter discharges produced by a radioactive material. Such counting is frequently accomplished by cascaded scale-of-two circuits similar to the one shown in Fig. 10·12, a sufficient number of circuits being employed to reduce the average pulse rate to the point where the counted down pulses can be recorded by an electromechanical counter. In radar work it is frequently necessary to reduce pulses produced by a crystal-controlled oscillator to a lower frequency for use as range marks (page 187) or as a system trigger.

It has been found that counting, or scaling, circuits based on blocking oscillators (page 159) give satisfactory operation at counting ratios as high as 1 to 10. Figure 10·18 gives the schematic diagram of a circuit of this type, together with representative waveforms (the steps in waveform at the test point are exaggerated

[6] J. L. Lawson, editor, Radiation Laboratory Report S10 May 1944; R. M. Ashby, F. W. Martin, and J. L. Lawson, Radiation Laboratory Report 914, March 1946; J. M. Sturtevant, Radiation Laboratory Report 654, Jan. 1945.

[7] E. C. Pollard and W. L. Davidson, *Applied Nuclear Physics*, John Wiley and Sons, Inc., Chapter 3, 1942.

in amplitude relative to the input pulses). The input pulses of constant amplitude, which may be spaced at irregular intervals, are taken from a low impedance source such as a blocking oscillator. The pulses are reduced in amplitude by the voltage divider composed of $C_1 < C_2$. C_3 is much larger than C_2 and therefore only affects the voltage division slightly; the purpose of C_3 is to furnish a low impedance test point at which the counting opera-

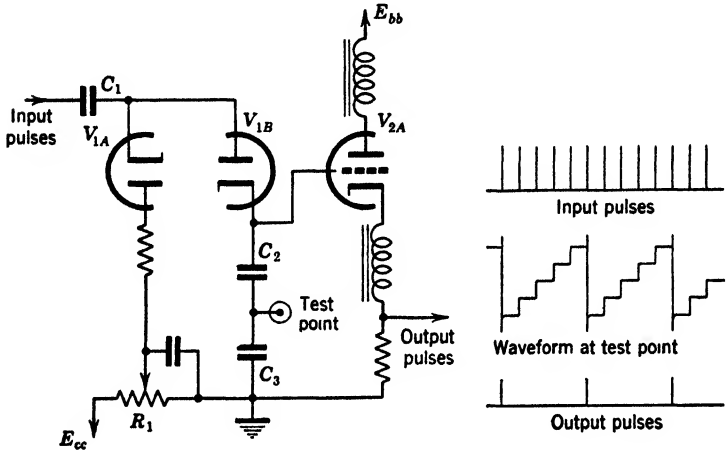


FIG. 10·18 Schematic diagram of a blocking oscillator counting circuit, and typical waveforms. This type of circuit can be employed with irregularly spaced pulses.

tion can be observed with a test oscilloscope without disturbing the circuit. The reduced pulses gradually charge C_2 along a step curve as shown until the potential of the grid of V_{2A} is increased to the point where the blocking oscillator fires. The heavy flow of grid current during the firing of the blocking oscillator rapidly discharges C_2 to a negative potential determined by the setting of R_1 and the value of E_{cc}. It is evident that the overall counting ratio is determined by several factors; (*a*) the size of the input pulses and the ratio of C_1 to C_2; (*b*) the negative bias supplied by R_1 and E_{cc}; (*c*) the cutoff potential of V_{2A}; and (*d*) to a lesser extent by the characteristics of the pulse transformer used. Stable counting by a factor of 5 can be obtained with no special precautions; with careful control of all voltages, including the filament voltage of V_{2A}, this factor can be extended to as high as 10. With

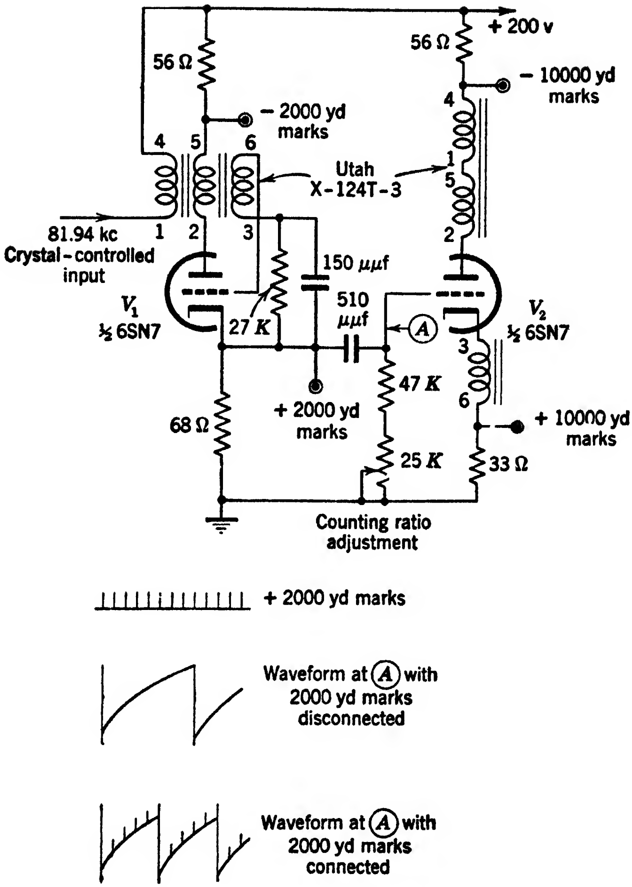


FIG. 10·19 Diagram of the blocking oscillator counting circuit used in the A and R scope (Radiation Laboratory), and representative waveforms. A circuit of this type may be used only with regularly spaced pulses.

all other conditions held constant, the counting factor can be varied over a small range by adjustment of R_1.

A somewhat simpler counting circuit is obtained by synchronizing a free-running blocking oscillator with pulses of higher frequency. It is evident that a scheme of this sort can be employed only with pulses of constant spacing. The circuit in Fig. 10·19, used in the so-called *A and R scope* developed at the Radiation

Laboratory, gives a stable counting ratio of 1 to 5 with 2000-yard range marks as the input. A qualitative understanding of the operation of this circuit can be obtained from the waveforms given in the figure. A detailed discussion is given by Easton and Odessey.[8]

10·6 GATED AMPLIFIERS

A gated amplifier is an amplifier which is operative only during intervals when a gating pulse, or gate, is applied to one of its electrodes. Several illustrations of the applications of gating have been given in earlier chapters. In a pentode, gating may be accomplished by applying a positive pulse to the plate, suppressor, screen, or control grid, or a negative pulse to the cathode, the gated electrode being normally biased so that no plate current flows except during the gating period. There is no need to discuss here the more or less obvious methods available for generating gating pulses and applying them to the gated tube.

If the amplifier to be gated has an appreciable response in the range of frequencies contained in the gating pulse, the output during the gate will have a different d-c level from that outside the gate because of the interrupted plate current. The output gated signals will thus ride on a so-called pedestal. For some purposes this pedestal is not objectionable, or may even be desirable. In other cases, as for example when the signals are to be applied to an intensity-modulated cathode ray tube which is not normally biased beyond visual cutoff, such a pedestal is not permissible. The formation of a pedestal may be prevented by adding a compensating tube which carries the plate current except during the gating period.

Figure 10·20 illustrates a non-pedestaling cathode-gated video stage which is convenient when the gating pulse is supplied by a 6SN7 multivibrator or Eccles-Jordan circuit. This stage has a gain of about unity with sufficient bandwidth to handle 1-microsecond pulses for most purposes, and is suitable for positive or negative signals of a few volts amplitude. V_{3B} is a compensating tube which carries the plate current when V_{3A} is gated off. V_{2A} and V_{2B} are cathode followers for driving the 6SL7 cathodes. When V_{1A} is conducting, the cathode of V_{3B} is held at a high

[8] A. Easton and P. H. Odessey, *Electronics*, **21**, 120 (May 1948).

enough potential so that this tube is cut off and, when V_{1B} is conducting, V_{3A} is cut off. Short switching transients will occur at the start and finish of each gating period; these are minimized by designing the source of the gating pulses to give as rapidly rising and falling gates as possible.

When the video signals are supplied from a low impedance source such as a 75-ohm cable, grid gating can be employed and

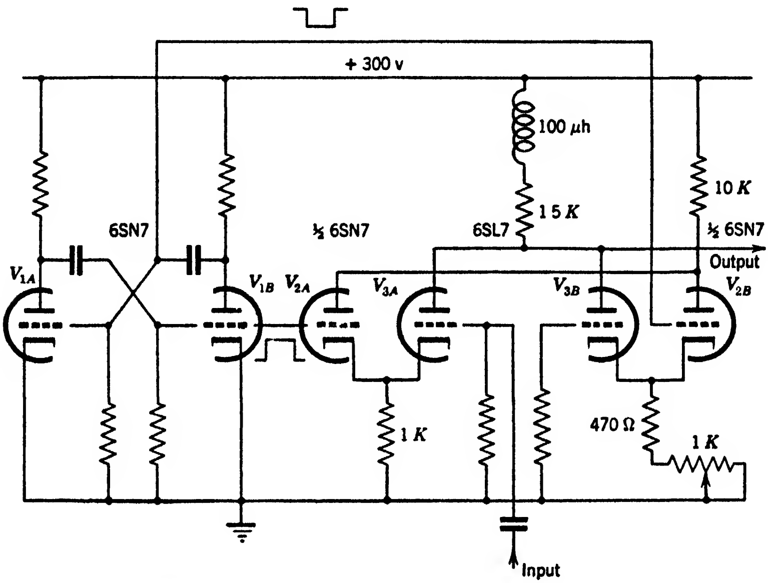


Fig. 10·20 Diagram of a non-pedestaling cathode-gated 6SL7 video stage. The gating pulses are supplied by the 6SN7 multivibrator. V_{3A} is the video stage, and V_{3B} is the compensating tube which carries the plate current while V_{3A} is cut off.

the signals applied to the cathode of the gated amplifier. If capacitive coupling is used, a large capacity will generally be required to give a sufficiently long time constant. The grids of the amplifier and compensating tubes are tied directly to the multivibrator grids. 6AC7's for high gain, or a 6SN7 for a gain of about unity, may be used for the amplifier and compensating tubes.

As an application of video gating we may cite the use of this type of circuit as a "local interference blanker" in both scientific and radar applications (cf. page 271). For example, video ampli-

fiers used in high speed and coincidence counting in nuclear research will probably be subject to interference from large video pulses radiated or conducted from high power nuclear machines such as linear accelerators. In such cases it will usually be far better to blank the counting circuit during each interfering pulse, thus discarding some small (measurable) fraction of the time, than to run the risk of getting spurious counts or coincidences. The blanking circuit will be synchronized with the interfering system by means of a trigger supplied by the latter.

PROBLEMS

10·1 Derive an expression for the minimum operating voltage drop in R (E_R in Fig. 10·1) necessary for certain striking of the arc in a VR tube, in terms of E_O, the operating voltage across the tube, E_S, the striking voltage, I_L, the operating load current, and I_{VR}, the operating tube current. Assume the load to be resistive. Calculate the minimum *supply* voltage needed in the case of a VR-105 (E_S = 133 volts, E_O = 104 volts), if I_L = 50 ma and I_{VR} = 20 ma.

10·2 Derive equations 10·1 and 10·2. Note that by definition

$$\mu_2 = \frac{d(E_b - E_0)}{d(E_0 - E_{G2})} \quad (I \text{ constant})$$

$$\frac{\mu_2}{r_{p2}} = \frac{dI}{d(E_0 - E_{G2})} \quad (E_b - E_0 \text{ constant})$$

$$\frac{1}{r_{p2}} = \frac{dI}{d(E_b - E_0)} \quad (E_0 - E_{G2} \text{ constant})$$

where the quantities have the significance shown in Fig. 10·4.

10·3 Draw the equivalent circuit of a cathode follower, and verify the results given in equations 10·3 and 10·4.

10·4 From a consideration of the equivalent circuit for the cathode-coupled amplifier of Fig. 10·7, derive the results given in Table 10·1. Calculate the approximate magnitude of the plate-to-plate a-c voltage for a 6SL7 cathode-coupled amplifier having $R_L = R_K = 100K$, with grid signals $e_1 = -e_2$ of 0.5-volt amplitude.

10·5 Show the derivation of equations 10·6 and 10·7. Calculate the expected frequency of a free-running multivibrator having the circuit constants given in Fig. 10·9.

10·6 Give the circuit constants for a counting circuit of the type shown in Fig. 10·18, which is to give a counting factor of 5 with 100-volt input pulses.

CHAPTER 11

RADAR AND ITS ACCESSORIES

Up to the present we have been considering the techniques which have evolved in the past few years under war pressure. In the three remaining chapters we propose to describe some applications of these techniques. There can be no doubt that for a combination of pulse circuitry, microwave technique, and modern cathode ray tube technique, radar is as good an illustration as one can choose. Moreover radar is not only a weapon which greatly helped to turn the tide of war; it has many peacetime uses, so that much more than historical interest attaches to it.

Three demands are made on radar: location, identification, and control. Radar is expected to detect the presence of some object, like a buoy in a harbor, or an aircraft approaching an airfield or a storm. It is expected to tell *which* buoy, what kind of aircraft, whether belonging to an airline or to the Army. There is also the demand to provide means by which something can be done about the object located, for example navigate a ship, or land an aircraft, or combine storm information with data from other stations to enable prediction of the weather.

These general demands assume different characters as soon as each is faced squarely. For example, the detection of buoys turns out to be very easy. The detection of high, fast, cylindrical rockets is not so easy, and the writers do not at present know of any satisfactory method of doing this. The marking of a buoy in a harbor is also very easy. The marking of aircraft is not very difficult, provided too much is not expected. The navigation of a ship is

very easy, but the control of dense air traffic near a large city needs more elaboration. Therefore different radars, unlike modern automobiles, really do look different. A ship navigation radar is small and simple; the same is true of an airborne navigation radar. A ground control radar is large and complex. A radar for landing purposes is even more specialized.

The impact of the demands on the radar designer accounts for the diversity of radars and explains the need for intensive research even after the basic problem has been solved.

The following paragraphs present a discussion of the physical basis of radar, attention being focused on pulsed microwave radar. This discussion is followed by brief descriptions of several radar sets representative of some of the important types developed for specialized applications. Finally some useful and interesting accessories of radars are considered.

11·1 PHYSICAL BASIS OF RADAR

It has been known for many decades that electromagnetic radiation incident on any discontinuity in the medium "carrying" the radiation is reflected to a greater or lesser extent. Obviously any material object such as those mentioned above furnishes such a discontinuity. It is the purpose of a radar system to radiate electromagnetic energy and to detect any of this energy which may be reflected from objects at some distance from the radar. At first thought it seems rather surprising that a small aircraft at a distance of 100 miles reflects enough energy to be detected. The amount of reflected energy which gets back to the radar is indeed extremely small, but it can nevertheless be well within the limits of detectability.

Obviously some means must be provided in the radar system for distinguishing the transmitted from the reflected energy. This is almost always accomplished by some system of *time* discrimination. The most widely used type of time discrimination is that employed in *pulsed* radars; our further discussion will be limited to radars of this type operating at frequencies in the microwave region. The radar transmitter is turned on for an interval of the order of a microsecond and thus radiates a short pulse of r-f energy. After an interval of time determined by the range of the

object to be detected, a reflected pulse of r-f energy returns and is suitably detected and interpreted. In this way the reflected energy is received at a time when there is no local disturbance to confuse the reception.

The transmitted and reflected energies can also be distinguished, with a continuous wave (c-w) radar, by means of the frequency modulation due to the Döppler effect produced by a moving target. Such radar may have future possibilities. but it was not in general use by the end of the war.

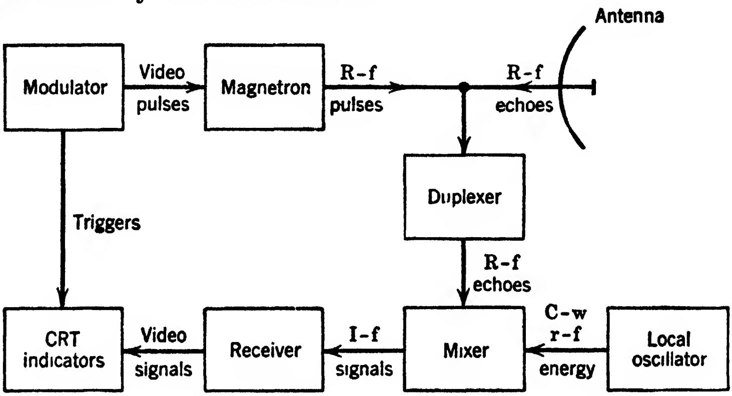


FIG. 11·1 Simplified block diagram of a pulsed radar system.

Figure 11·1 gives a highly simplified block diagram of a radar system. A modulator supplies large video pulses, at a repetition frequency of several hundred per second, to a magnetron or other microwave oscillator. The resulting r-f pulses are radiated by an antenna. The antenna in nearly all cases is rather highly directional, so that the radiated r-f power is confined to a restricted range of directions. Reflected power is picked up by the antenna and fed through a duplexer, which prevents damage to the mixer by the large r-f pulses from the transmitter, to the receiving system, where it is first converted to the intermediate frequency by the mixer and then greatly amplified and detected to give video signals. The latter are fed to appropriate indicators, which may number as many as 24 in high power search radars; the indicator sweeps are rigidly synchronized with the transmitted pulses by means of triggers obtained from the modulator.

The radiated pulse in its flight out to the target and back travels with the speed of light (186,284 miles per second). It is thus

necessary that the measurement of the time interval between the transmission of a pulse and the reception of the signal be carried out with a precision of the order of a microsecond or better if significant information about the range of the target is to be obtained. The data in Table 11·1 give an idea of the magnitude of the time intervals involved in radar ranging. It is interesting to note that, with some precision radars used for obtaining information for gun fire control, range is determined with an accuracy of about 20 yards, which corresponds to about 10^{-7} second.

TABLE 11·1 RANGES AND TWO-WAY TRANSMISSION INTERVALS

Range Unit	Range Units per Microsecond	Microseconds per Range Unit
Yards	163.9	0.00610
Statute miles	0.0932	10.73
Nautical miles	0.0809	12.36

When a directive antenna is used, means have to be provided for pointing the antenna in any desired position, or for causing it to sweep systematically all important directions, a process called *scanning*. Since it is important to know the direction of a target as well as its range, indicators such as PPI's (page 206) must be properly synchronized with the antenna position. This synchronization is accomplished by methods described in previous chapters. It can be readily appreciated that accurate pointing of an antenna, which may weigh as much as several hundred pounds, is a complicated task, especially if the ultimate support for the antenna is a rolling, pitching, yawing ship or aircraft.

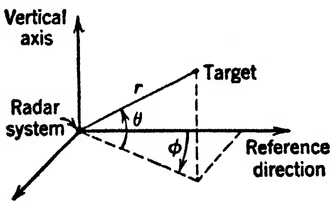


FIG. 11·2 Polar coordinate system customarily used in radar location; r = range; ϕ = azimuth; θ = elevation.

The coordinate system customarily used in radar practice is shown in Fig. 11·2. The direction of the radius vector from the radar to the target is specified by two angles. The azimuth ϕ is the direction of the projection of the vector on the horizontal from a fixed reference direction, usually taken as true north, and the elevation θ is the angle between this projection and the vector.

The Range Equation (Free Space Propagation)

It is important to consider the problem of how the maximum range at which a radar can detect a target is related to the various characteristics of the radar and the target. We will for the present assume that both the radar and the target are located in free space, so that interference and diffraction due to the earth's surface, and absorption and other effects due to the atmosphere, are absent.

Suppose that rectangular r-f pulses of peak power P_t are produced by the radar transmitter. The antenna concentrates this power into a beam so that the power radiated in the direction of *maximum* intensity is greater than P_t by a factor G called the antenna gain (Chapter 4, page 123). After the pulse has traveled a distance r in free space, the power density (power per unit area of the spherical wavefront) is equal to P_tG divided by the area $4\pi r^2$ of a sphere of radius r. If the pulse hits a target of *scattering cross section* σ (see page 339), the scattered power is the power intercepted by the area σ, which is

$$\frac{P_tG}{4\pi r^2}\sigma$$

σ is defined in such a way that this power is to be considered as isotropically scattered. After traveling the distance back to the radar, the pulse is again reduced in power by the "space attenuation" factor of $1/(4\pi r^2)$. If the *effective aperture* of the antenna (actual aperture multiplied by an efficiency factor ϵ having a value in the neighborhood of 0.6 for microwave antennas) is A, the peak power P_r delivered to the receiving system is

$$P_r = P_t \frac{GA}{(4\pi r^2)^2}\sigma \tag{11·1}$$

For the important case of an antenna consisting of a dipole radiator located at the focus of a parabolic reflector, G and A are related by the expression (cf. page 126)

$$A = \frac{G\lambda^2}{4\pi} \tag{11·2}$$

In this case we then have

$$P_r = P_t \frac{G^2\lambda^2}{(4\pi)^3 r^4}\sigma \tag{11·3}$$

It is convenient to express power ratios in decibels. If we define

$$
\begin{aligned}
db &= 10 \log \frac{P_r}{P_t} \\
db_g &= 10 \log \frac{G^2\lambda^2}{16\pi^2} \\
db_s &= 10 \log \frac{\sigma}{4\pi} \\
db_p{}^0 &= 10 \log \frac{1}{r^2}
\end{aligned}
\tag{11·4}
$$

equation 11·3 becomes

$$db = db_g + db_s + 2\, db_p{}^0 \tag{11·5}$$

Evidently, $db_p{}^0$ expresses the attenuation due to one-way transmission in free space, and has the values[1] given in Fig. 11·3. It may be noted here that equation 11·5 may be generalized to include conditions of propagation other than free space conditions by replacing $db_p{}^0$ by db_p, the field intensity at a point in *real* space referred to that at a point in *free* space 1 meter from the antenna. As we shall see later, db_p may be a complicated function of range and other factors.

In order to illustrate the orders of magnitude involved in equation 11·3 we may consider the case of the MEW (Microwave Early Warning), a high power ground-based radar which played an extremely important role in the war. This set operates at a wavelength of 10 centimeters, and develops a peak power of approximately 10^6 watts. The antenna reflector is a parabolic cylinder 24 feet long and 8 feet high, the gain being slightly less than 10^4 ($db_g = 38$). For a medium aircraft target having $\sigma = 1$ square meter ($db_s = -11$) at a range of 100 miles ($db_p{}^0 = -104$), we have

$$db = 38 - 11 - 208 = -181$$

so that

$$P_r \approx 10^{-18} P_t = 10^{-12} \text{ watt}$$

[1] The unit of length employed in calculating the values of $db_p{}^0$ in Fig. 11·3 is the meter, although they are plotted against statute miles. Thus the plot shows the field intensity at a point in free space relative to that at a distance of 1 meter from the antenna. Obviously the meter is to be used as the unit of length in expressing λ, σ, and A.

This is an extremely small amount of power, but as we shall see presently it is nevertheless readily detectable.

In Chapter 8 we discussed at some length the limitation imposed by *noise* on the sensitivity of a receiver. (In the present discussion the mixer is for convenience included in the receiver.) If we assume that the power $P_{\min}$ of the minimum detectable signal

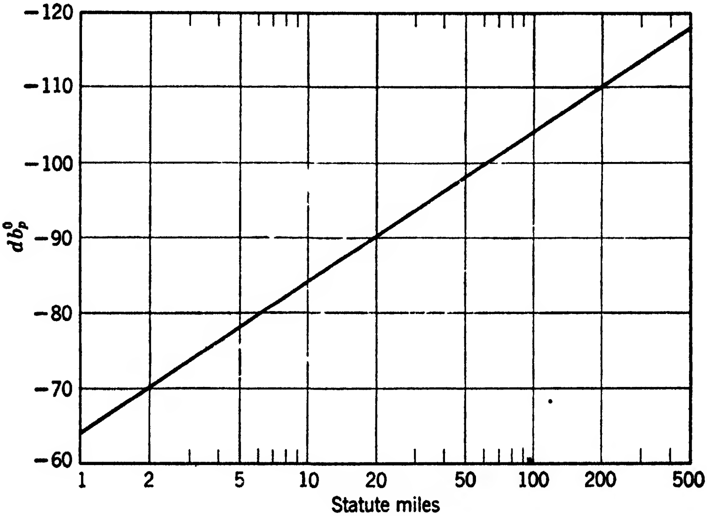


FIG. 11·3 Field intensity of a point in free space relative to the field intensity at a distance of 1 meter from an antenna, plotted as a function of the distance of the point from the antenna.

is equal to the effective noise power at the receiver input terminals,[2] then

$$P_{\min} = (NF)kT\,\Delta f \qquad (11\cdot6)$$

where (NF) is the receiver noise figure (page 247), k is Boltzmann's constant, T is the absolute temperature, and Δf is the bandwidth of the receiver. The bandwidth is usually determined primarily by the bandwidth of the i. f. amplifier (page 227). The *maximum*

[2] This assumption is an approximation. Signals having powers less than the receiver noise power by several decibels can be detected under some circumstances, because the signal pulses have different shapes from noise pulses. On the other hand, factors such as scanning losses may result in the minimum detectable signal power being considerably larger than the receiver noise power. These points are given further attention below.

range $r_{\max}$ at which a target of scattering cross section σ can be detected is thus given by the *range equation*

$$r_{\max} = \left(\frac{P_t}{P_{\min}} \frac{G^2\lambda^2}{(4\pi)^3}\sigma\right)^{1/4} \tag{11·7}$$

It is to be remembered that this equation applies only to a target in the direction of the maximum of the antenna pattern. Since the value of $r_{\max}$ depends explicitly on the extra-radar factor σ, as well as implicitly on propagation, a more significant measure of radar performance is the *performance figure:*

$$db_0 = 10 \log \frac{P_t}{P_{\min}} \frac{G^2\lambda^2}{(4\pi)^2} \tag{11·8}$$

It is evident that the range equation can be written

$$-2(db_p)_{\max} = db_0 + db_s \tag{11·9}$$

The MEW receiver has a noise figure of approximately 10 decibels and a bandwidth of about 1.6 megacycles; therefore $P_{\min} \approx 6.4 \times 10^{-14}$ watt ($T = 300°K$), and $db_0 = 230$. Taking again the value $db_s = -11$ (medium aircraft target), $(db_p)_{\max} = -109.5$. *If free space conditions apply,* reference to Fig. 11·3 shows that the MEW should be able to see a medium aircraft at a maximum range of about 180 miles (if no allowance is made for scanning losses).

Equation 11·7 brings out the fact that $r_{\max}$ increases very slowly with increases in transmitted power or receiver sensitivity. If P_t is doubled, $r_{\max}$ is increased by only 19 per cent.

11·2 DISCUSSION OF THE RANGE EQUATION

There are several inter-relations between the various terms in equation 11·7 which are not immediately obvious. It is the purpose of this section to discuss some of these relations. Two types of factors need to be considered, namely those characteristic of the radar itself and those independent of the radar.

Antenna Gain and Beam Width

The quantities G and λ appearing in the range equation are not independent. For a parabolic reflector of circular aperture, equation 11·2 and the definition of the efficiency factor ϵ show that

$$G = \epsilon\pi^2 \frac{D^2}{\lambda^2}$$

where D is the diameter of the reflector. Substitution in equation 11·7 gives

$$r_{\text{max}} = \left[\frac{P_t}{P_{\text{min}}} \frac{\pi\epsilon^2}{64} \left(\frac{D}{\lambda}\right)^4 \lambda^2\sigma\right]^{1/4}$$

The beamwidth is roughly proportional to the ratio λ/D; thus r_{max} is increased by an increase in wavelength if the reflector size is also increased to maintain constant beamwidth. However, r_{max} is more strongly dependent on beamwidth than it is on wavelength, so that it is generally advantageous, within limitations imposed by such diverse factors as scanning losses and available oscillator powers, to use the narrow beamwidths obtainable with antennas of reasonable size in the microwave region, rather than to go to longer wavelengths and broader beams. Narrow beamwidth gives the added advantage of improved azimuth and elevation discrimination between targets.

Receiver Bandwidth

Equation 11·6 shows that the power of the minimum detectable signal is directly proportional to the receiver bandwidth, so that it is important to make the latter no greater than necessary. However, as the bandwidth is decreased, the rise time of the output pulse is increased because of attenuation of the high frequency components in the pulse spectrum (cf. Fig. 5·3), and a point will eventually be reached at which the output pulse does not have time to achieve its proper full amplitude in the short interval corresponding to the input pulse width. Actually no definite statement can be made about the best compromise which can be reached here, since the optimum bandwidth depends on a variety of factors, such as how the output pulses are going to be displayed and what uses one wishes to make of the displayed pulses. J. L.

Lawson and his group at the Radiation Laboratory have made extensive experiments along this line. They have found, for example, that for the A-scope (Chapter Six) type of presentation the overall sensitivity of signal detection is greatest when the receiver bandwidth is about $1.2/\tau$, where τ is the input pulse width. The minimum discernible signal under these conditions is increased by 0.6 decibel by doubling or halving the optimum bandwidth, and by about 5 decibels if the bandwidth is changed by a factor of 10 or 0.1.

In some cases minimizing $P_{\min}$ may be a matter of secondary importance in selecting the receiver bandwidth. Thus, a wide-band receiver may be selected because a short pulse rise time is desired for purposes of accurate range measurements, or in order to decrease the difficulty of receiving a signal transmitted by a beacon (Section 11·4) which may not be tuned exactly to the radar signal, or even to provide greater leeway in tuning the receiver to its own transmitter.

Pulse Width and Average Power

The average power P_{av} radiated by a transmitter is equal to the peak power multiplied by the *duty cycle* δ, defined as the fraction of the time the transmitter is actually radiating. Evidently the duty cycle is given by

$$\delta = \tau f_r \tag{11·10}$$

where f_r is the pulse repetition frequency. In the case of the MEW, $\tau = 1$ microsecond and $f_r = 350$, so that $\delta = 3.5 \times 10^{-4}$ and $P_{av} = 350$ watts. The MEW is considered to be a high power set, even though its average power is only 350 watts.

As pointed out above, in many radar applications the optimum receiver bandwidth is reciprocally related to the pulse width:

$$\Delta f = \frac{m}{\tau}$$

where m is a factor usually not very different from 1·2. Substitution of equations 11·6 and 11·10 in the range equation (equation 11·7) gives

$$r_{\max} = \left(\frac{P_{av}}{f_r}\frac{1}{m(NF)kT}\frac{G^2\lambda^2}{(4\pi)^3}\sigma\right)^{1/4} \tag{11·11}$$

It is thus evident that, *provided the receiver bandwidth is held at its optimum value* (m = constant), the maximum range depends on the *average* rather than the *peak* power. So far as range alone is concerned, there is no real merit in using very short bursts of very high power, because this requires the use of a wideband receiver with accompanying large noise power. However, there are two generally valid reasons for using the available average power in the form of short pulses. In the first place, range resolution is better with short pulses; two targets at approximately the same azimuth and elevation cannot be distinguished if they are closer together in range than the distance corresponding to approximately one pulse width. A second consideration favoring short pulses is that the signal energy returned from an aircraft or other relatively small target decreases more slowly as the pulse width decreases than does the return from clouds, sloping land, and jamming devices such as "window" (page 269), because the energy in a single reflected pulse can arise from reflections occurring at points separated in range by a distance of the order of one pulse width. This integration effect means that short pulses favor the type of target of usual interest as compared with targets considered in most cases as interfering.

Repetition Frequency

According to equation 11·11, if *peak* power and all other factors are held constant, doubling the repetition frequency will have no effect on r_{max}, since the ratio P_{av}/f_r will remain constant. However, two points which influence the choice of f_r may be mentioned. Because of the integrating effect of the fluorescent screens used in radar indicators (Chapter 6) and of the human eye, the minimum discernible signal is somewhat decreased by an increase in f_r. It has been found empirically that P_{min} is approximately proportional to $f_r^{-1/2}$. On the other hand, f_r cannot be increased too much, since time must be allowed between transmitted pulses for echoes to return from the most distant targets the radar is expected to detect. That is, f_r must be less than the velocity of light divided by twice the maximum expected range.

In view of the preceding comments, the question arises as to whether a given available average power is most effectively utilized by increasing the repetition rate or the peak power. Since the

signal discernibility increases roughly as the square root of f_r and as the first power of P_t, the answer is obviously to increase P_t, provided the equipment permits this to be done.

Pulse Shape

It has been assumed throughout the foregoing discussion that the transmitted pulse envelope approximates closely to a rectangular shape. It is true that rectangular pulses contain considerable power in sideband components too far removed from the carrier frequency to be passed by a receiver of optimum bandwidth, so that one might expect a more efficient utilization of power would be obtained by using somewhat rounded pulses. However, high frequency oscillators undergo considerable frequency modulation when they are amplitude-modulated, so that if the modulating pulse amplitude rose slowly, the output r-f frequency would not come within the receiver pass band until nearly full pulse amplitude had been reached.

Scanning and Scanning Losses

In order to cover the solid angle of interest it is necessary to move the antenna beam in some more or less systematic way known as scanning. Several types of scanning are employed, the commonest being circular scanning, in which the antenna is rotated continuously about a vertical axis, and sector scanning, in which the antenna is rotated back and forth through a certain angle about a vertical axis. Circular scanning is ordinarily employed with search radars such as the MEW, adequate vertical coverage being obtained by some such device as having the antenna pattern broader in elevation than in azimuth.

It is evident that as the scanning rate is increased there will eventually be an appreciable probability that no pulse will hit the target during the very brief interval when the beam is traversing the target. Long before this extreme condition is reached there is an appreciable loss in effective signal strength, which must be carefully considered in designing a radar. Lawson and his coworkers have studied this problem in detail and have arrived at the following semiempirical formulation of scanning losses. If T_F is the period of the scanning motion in seconds, F is the frac-

tion of the time the antenna beam is on the target, and S_{90}/N is the ratio of the signal power required for visibility 90 per cent of the time (under carefully specified test conditions) to the noise power, then

$$\frac{\dfrac{S_{90}}{N}}{\left(\dfrac{S_{90}}{N}\right)_0} = \sqrt{\frac{8}{FT_F}} \quad \text{if } T_F > 8 \text{ seconds} \tag{11·12}$$

$$= \sqrt{\frac{1}{F}} \quad \text{if } T_F < 8 \text{ seconds} \tag{11·13}$$

Here $(S_{90}/N)_0$ is the signal to-noise ratio under "searchlighting" conditions ($F = 1$); therefore the ratio on the left side of these equations can be interpreted as the reciprocal of the scanning loss. Equation 11·12 shows that the scanning loss becomes unity if the scan rate is decreased until it takes 8 seconds for the antenna beam to traverse the target. The loss is greater than indicated by equations 11·12 and 11·13 for very high scan rates or very narrow beams. It is somewhat less than predicted for stationary targets since then advantage is gained from the integrating effect of the persistence of the indicator screen. The improvement from this cause is at most a little more than 1 decibel.

The scan rate of the MEW is 4 revolutions per minute, so that $T_F = 15$. The beam width in azimuth is 1°, so that $F = 1/360$ for small targets such as aircraft. Therefore the scanning loss is 1/13.8, or 14 decibels. The loss would be increased by 3 decibels if the scan rate were increased to 16 revolutions per minute, or the beam width decreased to 0.25°. It should, however, be noted that a decrease in beam width caused by an increase in antenna size would be accompanied by an increase in antenna gain which would far outweigh the increased scanning loss, provided the beam width is not below the range of applicability of equations 11·12 and 11·13. Increase in *one* dimension of an antenna of rectangular aperture causes an approximately proportionate increase in gain; therefore r_{max} is proportional to the square root of antenna dimension so far as this effect is concerned, and is inversely proportional only to the one-eighth power of the dimension so far as the scanning loss is concerned, leaving a net effect proportional to the three-eighths power of the dimension.

A further limitation, apart from the scanning loss discussed above, on the permissible scanning speed is imposed by the finite velocity of the radar pulse. It is of course necessary that the reflected pulse return to the radar antenna before this has turned so far that it cannot detect the returning pulse. If we consider, for simplification, an antenna which scans in azimuth only, the rate of antenna motion in degrees per second should be considerably less than the beam width in degrees divided by $10.7 r_{max} \times 10^{-6}$, where r_{max} is the maximum range, in miles, at which target detection is to be accomplished. Thus, a radar with a 1° beam used for detection out to 100 miles should have a scan rate less than 3 revolutions per second.

Extra-Radar Factors Affecting Performance

Two factors external to the radar which affect r_{max} are of fundamental importance: the propagation of the transmitted and reflected power through the space between the radar and the target, and the scattering of the transmitted power by the target. Both of these subjects are quite complex, and a complete discussion of them would be outside the scope of this book. We shall merely mention briefly some of the more important points and attempt to give some idea of the magnitudes of the effects commonly encountered.

Propagation

The assumption of free space propagation which has been made so far in this chapter obviously cannot be exact in radars functioning on or near the surface of the earth, although it is a remarkably good approximation in microwave radars looking over land of average roughness at airborne targets not too near the horizon.

Electromagnetic radiation at very high frequencies is restricted to propagation over essentially line-of-sight paths, except for diffraction effects which permit seeing around small objects or over the horizon to a small extent, and for occasional periods of abnormal atmospheric refraction. Thus a land-based radar which is used for detecting aircraft should be sited so that no hills cause

shielding above the minimum elevation angle of importance.[3] The most obvious result of line-of-sight propagation is the existence of a radar horizon. As a general rule the atmosphere nearest the earth's surface has a very slightly higher index of refraction, which decreases approximately linearly with altitude, for radiation in the range of radar frequencies, so that the radar horizon is a little more distant than the optical horizon. This refractive effect is usually accounted for by assigning an effective radius to the earth about 1.33 times its actual radius. Thus, for an antenna height h_a and target height h_t, the horizon range r_h is

$$r_h = \sqrt{\frac{8}{3}a}\,(\sqrt{h_a} + \sqrt{h_t}) \tag{11·14}$$

where a is the earth's actual radius. It happens that if r_h is expressed in statute miles and h_a, h_t in feet, this equation becomes very nearly

$$r_h = \sqrt{2}(\sqrt{h_a} + \sqrt{h_t}) \tag{11·15}$$

It sometimes happens that the rate of decrease of refractive index with altitude becomes abnormally large, as, for example, when hot dry air from a land mass moves out over cold water. This leads to the formation of so-called trapping layers[4] such that the radiation from a source below the top of the layer extends with roughly inverse square law attenuation to distances quite far beyond the normal horizon at low altitudes, but with high attenuation at high altitudes. Under such conditions the radar range for targets close to the surface may be several times the normal value, while high-flying aircraft are not detected beyond very short ranges. As reported by Pekeris,[4] a rapid decrease of moisture with elevation commonly occurring at Antigua in the British West Indies leads to persistent ducts 20 to 40 feet high, as a result of which 10- and 3-centimeter radiation from antennas close to the surface is propagated out to ranges as great as 200 miles.

[3] Ground radars are sometimes deliberately sited with *low angle* shielding close to the radar to reduce the confusing display of echoes from fixed land targets on the radar indicators. It is evident that under normal radar operation an aircraft target at any elevation would not show up if it were at the same range and azimuth as a land mass giving a strong area of signals on the indicators.

[4] C. L. Pekeris, *Proc. I.R.E.*, **35**, 453 (1947).

At very high frequencies, absorption of electromagnetic radiation, chiefly by water in liquid form, may very seriously affect radar performance. The estimated figures[5] given in Table 11·2 are indicative of orders of magnitude to be expected.

TABLE 11·2 ESTIMATED ABSORPTION OF RADIATION BY RAINFALL

Description of Rainfall	Precipitation (mm per hr)	Attenuation (db per mile)		
		3,000 mc	10,000 mc	30,000 mc
Fog	0.01	.....	.....	0.0033
Drizzle	0.1	.....	0.005	0.033
Light rain	1	0.004	0.05	0.33
Heavy rain	10	0.04	0.5	3.3
Cloudburst	100	0.4	5	33

It should be emphasized that trouble may also be experienced as a result of radiation *scattered* rather than absorbed by areas of dense clouds or rainfall, because of the strong signals returned from such areas. Thus it is frequently observed that aircraft signals are completely lost when they coincide in range with strong cloud signals, although aircraft beyond the cloud mass are clearly visible, showing that serious *absorption* has not taken place.

Propagation may be markedly affected by interference effects resulting from reflection from the earth's surface. Suppose we have an isotropic radiator A (Fig. 11·4) at a height h_a above a flat, perfectly reflecting earth, and a target B at a height h_t and range r_s. The effect of the reflecting surface is the same as though the reflector were removed and an *image radiator* placed at D with its radiation equal in magnitude but opposite in phase to that from the primary radiator. The difference in the path lengths r_s and $\rho_1 + \rho_2$ causes an additional phase shift β at the target:

$$\beta = \frac{2\pi}{\lambda} \Delta r$$

[5] S. C. Hight, Bell Laboratories Report MM-44-170-50, Oct. 1944. See also S. D. Robertson and A. P. King, *Proc. I.R.E.*, **34**, 178P (1946); G. E. Mueller, *ibid.*, **34**, 181P (1946); L. J. Anderson, J. P. Day, C. H. Freres, and A. P. D. Stokes, *ibid.*, **35**, 351 (1947).

where $\Delta r = \rho_1 + \rho_2 - r_s$. If E_0 is the free space amplitude at the target, the actual amplitude E is

$$E = E_0\sqrt{2(1 - \cos\beta)} = 2E_0 \sin\frac{\pi}{\lambda}\Delta r$$

provided Δr is small enough so that any additional space attenuation in the reflected energy can be neglected. If the target is small enough so that it can be assumed to be illuminated by radiation of

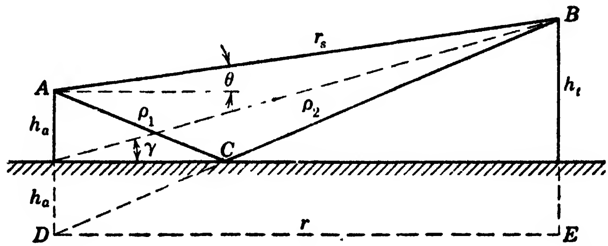


FIG. 11·4 Diagram illustrating the effect of a flat reflecting surface on the radiation from an isotropic radiator. The intensity at B from the radiator at A is the vector sum of the intensities due to transmission over the direct path AB and the indirect path ACB.

constant amplitude and phase, the power density at the target is equal to the free space power density multiplied by the factor

$$F^2 = 4\sin^2\left(\frac{\pi}{\lambda}\Delta r\right)$$

Since the radar case involves two-way transmission, it is seen that as h_a, h_t, or r_s varies, the power received by scattering from the target will oscillate between zero and sixteen times the free space value; therefore, according to this idealized picture, one might expect a maximum range as much as twice the free space value, obtained, of course, at the expense of regions where no echo is detected at all. This assumes that the pattern of radiation from the antenna at A is broad enough so that the intensity of the reflected energy is equal to that of the direct energy, which is true in most cases of interest.

Reference to Fig. 11·4 shows that

$$(\rho_1 + \rho_2)^2 = r_s^2 - (h_t - h_a)^2 + (h_t + h_a)^2 = r_s^2 + 4h_ah_t$$

so that

$$\Delta r = r_s\left(1 + \frac{4h_a h_t}{r_s^{\,2}}\right)^{1/2} - r_s \approx \frac{2h_a h_t}{r}$$

This approximation is valid to within 1 per cent for values of $\gamma \approx \theta$ less than 8°. Since for these small angles $\sin\theta \approx \theta$

$$\Delta r \approx 2h_a\theta$$

and

$$F^2 \approx 4\sin^2\left(\frac{2\pi}{\lambda} h_a\theta\right) \tag{11·16}$$

In terms of db_p (page 322)

$$db_p = db_p{}^0 + 20\log F \tag{11·17}$$

The field intensity pattern from the radiator at A thus has maxima at

$$\theta = \frac{2n+1}{4}\frac{\lambda}{h_a}, \quad n = 0, 1, 2, \cdots \tag{11·18}$$

and nulls at

$$\theta = \frac{n}{2}\frac{\lambda}{h_a}, \quad n = 0, 1, 2, \cdots \tag{11·19}$$

for small elevation angles θ.

The radiation pattern is broken up into lobes as shown in Fig. 11·5, in which are plotted the contours of constant field intensity equal to 98 decibels below the value 1 meter in front of the antenna; that is, the contours of $db_p = -98$, for the case of radiation at 3000 megacycles ($\lambda = 10$ centimeters) from an antenna 50 feet above a flat perfectly reflecting earth. Several results of this radiation pattern may be noted:

(*a*) The maximum range at which a given field intensity is obtained is twice the free space range.

(*b*) The field intensity is low near the surface of the earth; therefore a target approaching the antenna at low elevation would not be detected until short range.

(*c*) There are more or less extensive regions of low intensity in the vicinity of the null angles; therefore a target moving in at constant altitude would go through a succession of "fading" regions.

(*d*) The elevation angle of the first lobe and the angular separation of the lobes are directly proportional to the wavelength and inversely proportional to the antenna height. Thus radars operating at very high frequencies are better adapted than those at low frequencies to detecting low-flying aircraft, or surface targets, and give more solid coverage in that the fading regions are narrower. Conversely, such radars do not require as large antenna heights

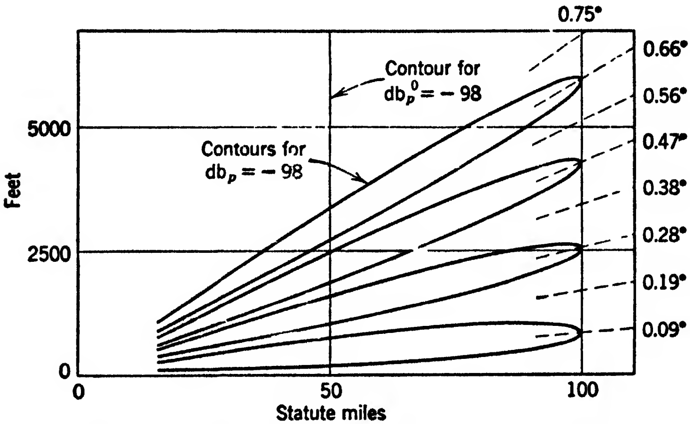


FIG. 11·5 Plot of the ranges and altitudes at which the field intensity from a 3000-mc radiator located 50 ft above a flat, perfectly conducting earth is 98 db less than the intensity 1 meter from the radiator.

to give a specified low angle coverage. This small lobe spacing and low angle coverage constitute the greatest advantage of microwave radar in shipboard applications.

The flat earth approximation gives useful results under some conditions. However, a variety of complicating factors must be taken into consideration in many cases. These factors may be listed as follows:

(*a*) The earth's curvature introduces three complications. We have already discussed the simple horizon effect. A second effect results because rays reflected from a convex surface are more divergent after reflection than before. The third effect is diffraction, as a result of which there is a finite field intensity at points below the horizon.

(*b*) The reflection coefficient of land and water is not usually equal to -1 (magnitude unity and phase angle π). It is actually a

function of the roughness, dielectric constant, and conductivity of the surface, and of wavelength, polarization, and angle of incidence of the radiation. With regard to roughness, an idea of what may be expected can be obtained from Rayleigh's criterion that irregularities should disturb the phase by less than $\pi/4$ radians. Thus the height of irregularities in the surface should be less than about $\lambda/16\alpha$, where α is the angle of incidence, in radians, measured from the horizontal. Experience shows that ordinary land is sufficiently rough that microwaves undergo negligible reflection from it, in the region of grazing incidence, while radiation at 100 to 200 megacycles is reflected very well if the terrain is reasonably smooth and cleared of trees and bushes. Thus microwave radars sited inland generally behave as predicted on the basis of free space propagation, whereas long wave radars have to be sited with great care in order for their behavior to be predictable. There is very little practical experience to show how interference resulting from water surfaces varies with surface roughness. However, interference even with microwaves shows up under a wide range of weather conditions. The remaining factors listed above as affecting reflectivity are interrelated in a complicated way. So far as sea water is concerned, the reflection coefficient is very nearly -1 at grazing angles up to several degrees for frequencies up to microwave frequencies for horizontal polarization, while for vertical polarization the magnitude and phase angle of the coefficient both decrease quite rapidly from the initial values of 1 and π as the angle of incidence increases. In cases where it is desired to use interference patterns, horizontal polarization will always be employed.

Figures 11·6 and 11·7 show the results of calculations which take these effects into consideration. Figure 11·6 is a plot of the quantity $db_p - db_p{}^0$ as a function of range for an antenna height of 30 feet and a constant altitude of 500 feet for propagation of 3000-megacycle radiation over sea water with both vertical and horizontal polarization. It is seen that the greater reflection with horizontal polarization leads to a more pronounced lobe structure; that the last rise above the free space value occurs somewhat inside the "horizon" as defined by Equation 11·15; and that there is a very rapid fall just inside and beyond the horizon in the diffraction region. This fall corresponds to increasingly large nega-

tive powers of r in equation 11·3. Figure 11·7 shows the first interference maxima for 3000- and 100-megacycle radiation over

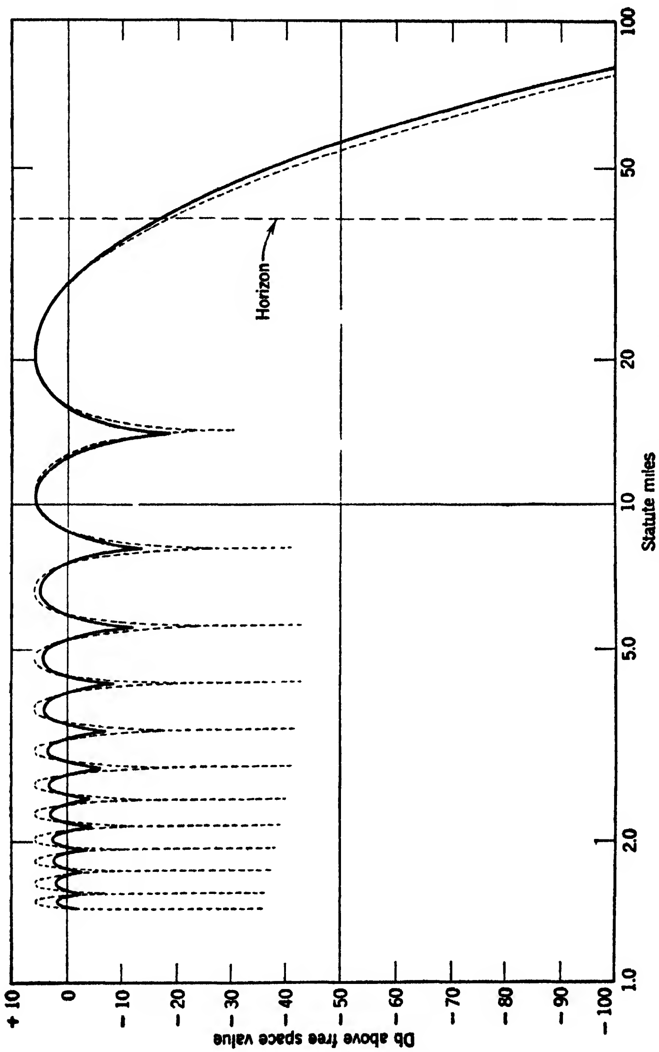


FIG. 11·6 Calculated values of the field intensity at an altitude of 500 ft relative to that in free space ($db_p - db_p^0$) for a 3000-mc transmitter located 30 ft above the sea. The solid curve is for vertical polarization and the dashed curve for horizontal polarization. (Effective earth's radius = 1.33 × actual radius.)

sea water, computed for an antenna height of 500 feet, horizontal polarization being used. The great superiority of the higher frequency for low angle coverage is evident.

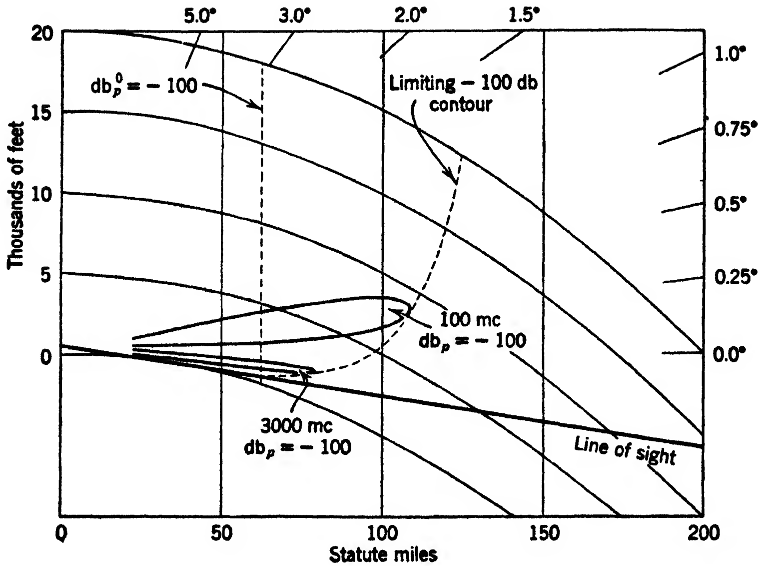


FIG. 11·7 Calculated first interference maxima for the propagation of 3000-mc and 100-mc radiation over sea water from an antenna height of 500 ft. (Horizontal polarization; effective earth's radius = 1.33 × actual radius.)

Scattering

The scattering of energy by targets of very simple shape can be calculated with sufficient accuracy for most microwave radar applications by the methods of geometrical optics.[6] Consider, for example, the case of a spherical surface of perfectly reflecting material illuminated with parallel radiation (Fig. 11·8). The energy reflected from the sphere appears to come from a focus $R/2$ in back of the surface, R being the radius of the sphere. If A is the (circular) aperture of the receiving antenna located at a range r from the sphere along the line of the incident radiation, the area a which defines the cylinder of incident radiation which will be intercepted by the antenna after reflection is given by

$$a = \frac{AR^2}{4r^2}$$

[6] J. F. Carlson and S. A. Goudsmit, Radiation Laboratory Report 195, Feb. 1943; R. C. Spencer, Radiation Laboratory Report 661, Jan. 1945.

It was shown on page 321 that the power density at the reflector is $P_tG/4\pi r^2$. It follows that the power received at the radar is

$$P_r = \frac{P_tG}{4\pi r^2}\frac{AR^2}{4r^2} \tag{11·20}$$

Comparison of this equation with equation 11·1 shows that for a reflecting spherical segment the *scattering cross section* is

$$\sigma = \pi R^2 \tag{11·21}$$

a quantity equal to one-fourth the area of the corresponding sphere. The radius of curvature is the determining quantity, and the cross section is independent of the wavelength (provided the spherical

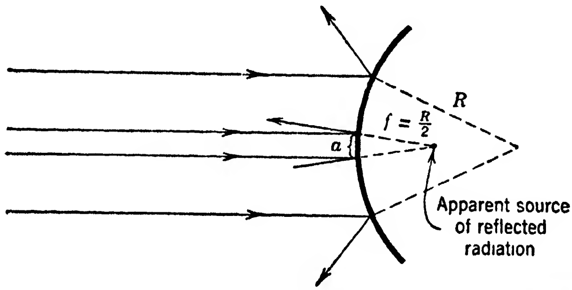


FIG. 11·8 Reflection of parallel radiation by a perfectly reflecting material. Only the radiation incident on a small area a will be intercepted by the receiving antenna after reflection.

segment is not too small). It is easily seen that the same result is obtained if a concave spherical surface is illuminated.

The calculation of σ values becomes rapidly more involved for more complicated cases. We list here a few results,[6] again for perfectly reflecting, uniformly illuminated surfaces.

(*a*) Flat surface of area S perpendicular to the radar beam (cf. equation 11·2):

$$\sigma = \frac{4\pi S^2}{\lambda^2} \tag{11·22}$$

(*b*) Flat surface the normal to which makes an angle ξ with the radar beam:

$$\sigma \approx \frac{4\pi\lambda^2}{(2\pi\xi)^4} \tag{11·23}$$

[6] J. F. Carlson and S. A. Goudsmit, Radiation Laboratory Report 195, Feb. 1943; R. C. Spencer, Radiation Laboratory Report 661, Jan. 1945.

(c) Cylinder of radius R and length l perpendicular to the radar beam:

$$\sigma = \frac{2\pi R l^2}{\lambda} \tag{11·24}$$

(d) Cylinder of radius R the normal to which makes an angle ξ with the radar beam:

$$\sigma \approx \frac{R\lambda}{2\pi\xi^2} \tag{11·25}$$

It is to be noted that these cross sections are all dependent on the wavelength. Equations 11·23 and 11·25 also show a strong dependence on the orientation of the reflecting surface.

A special type of reflector, known as a *corner reflector*, has received considerable attention because it has a scattering cross section comparable to the maximum cross section shown by a flat

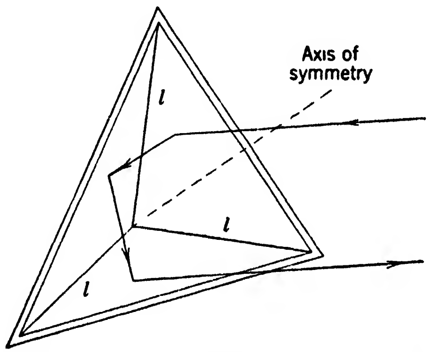


FIG. 11·9 A corner reflector with triangular sides. The arrows indicate an incident ray undergoing a triple reflection, and illustrate how the ray is finally reflected back in the direction of its origin. This property of corner reflectors results in their having large scattering cross sections with relatively weak dependence on orientation.

surface of similar size, but showing much less dependence on orientation. A corner reflector (Fig. 11·9) is composed of three mutually perpendicular planes and has the property that radiation incident upon it is reflected in the same direction provided at least two of the reflecting surfaces are visible from the direction of the incident radiation. In such a case the radiation undergoes either two or three reflections before leaving the reflector. A cor-

ner reflector with isosceles triangular sides has a maximum scattering cross section given by [7]

$$\sigma = \frac{4\pi l^4}{3\lambda^2} \tag{11·26}$$

where l is the length of the sides, provided l is considerably larger than λ. Thus, such a reflector having 1-foot sides has a maximum scattering cross section equal to 39 square feet at $\lambda = 10$ centimeters and 380 square feet at $\lambda = 3$ centimeters. Spheres of diameters 3.5 feet and 22 feet respectively would be required to give these same scattering cross sections.

Practical targets such as aircraft and ships are so complicated that it is hopeless to consider their scattering cross sections from any viewpoint other than the empirical. Any such target will give a radar return made up of numerous contributions of various phases from reflecting surfaces within the region defined by the beam width and pulse length of the radar. When it is considered that at $\lambda = 3$ centimeters a relative motion of two surfaces of less than 1 centimeter can change their contributions from constructive to destructive interference, it is not surprising that radar echoes frequently show violent fluctuations in intensity. H. Goldstein and P. Bales have found by photographic recording of individual echo pulses that the return from a large ship, for example, may show rapid fluctuation at least as large as 35 decibels occurring at rates as high as 1500 decibels per second. Very interesting studies of the echo strength from an airplane on the ground have been made by Ashby, Martin, and Lawson.[8] A 10-centimeter radar with very high repetition frequency (10,000 pulses per second) was used, and care was taken to minimize effects from reflections from the ground. The signal strength from the aircraft was automatically recorded on a polar graph as the aircraft was rotated. An extremely complex lobe pattern was obtained in this way, with intensity maxima at irregular intervals of the order of 0.5°, the change between minimum and maximum being of the order of 15 to 20 decibels! Broadside returns averaged 10 to 15 decibels higher than bow or stern returns.

[7] R. C. Spencer, Radiation Laboratory Report 433, March 1944.

[8] R. M. Ashby, F. W. Martin, and J. L. Lawson, Radiation Laboratory Report 914, March 1946.

The above considerations indicate that it is difficult to measure accurately the scattering cross sections of practical targets. Other factors, such as psychological factors involved in observations of signals on cathode ray tube indicators,[9] contribute further to this difficulty. Values in excess of 10^6 square meters have been estimated for large ships.

11·3 DESCRIPTION OF SOME RADAR SETS

By way of illustration of the above general account of radar we include short accounts of five sets together with the general reasons for the choice of the various design features. All but one of these sets are or have been in production. The exception is the experimental "Cindy" set, which is ideal for the important purpose of ship navigation.

Air Navigation

One of the important needs during the war was for an airborne radar to navigate strategic bombers through clouds and at night, and in case of necessity to enable them to attempt to hit a target without visual aids. Such a radar is the "H_2X" or AN/APS-15.[10] A lightweight version of this radar came into production at the end of the war and contains many design features of interest. This is the AN/APS-10. The design characteristics for this set are [11] as follows

Wavelength	3 cm
Output power	8 kw at 800 pulses per second
Pulse length	0.8 μsec
Antenna size	18-in. paraboloid, modified to give a cosecant squared beam
Antenna gain	700
Receiver sensitivity and bandwidth	15 db below $kT\,\Delta f$; $\Delta f = 6$ mc
Type of indicator	PPI
Weight	100 lb
Power consumption	350 watts
Rate of scan	15 rpm

[9] L. B. Linford, D. Williams, V. Josephson, and W. Woodcock Radiation Laboratory Report 353, Nov. 1942.

[10] J. V. Holdam, S. McGrath, and A. D. Cole, *Electronics*, p. 138, May 1946; p. 142, June 1946.

[11] J. V. Holdam, S. McGrath, and A. D. Cole. *Electronics*, p. 132, Feb. 1947.

This radar is the lightest microwave airborne radar which has been made. It embodies a low power magnetron which delivers less power than most production microwave magnetrons. This is no disadvantage because the function of the set is purely for navigation, and the only targets ever scanned are large land objects or shipping (which has a large scattering cross section). This fact reduces the need for high power and makes possible the construction of a set which is far lighter than sets which have to detect aircraft or submarines as part of their functions. The design of the set has the reduction of weight as the prime objective. Second to this is reliability of performance without tuning in the air. This last factor is the reason for the unusually large bandwidth which is not called for by a pulse width of approximately 1 microsecond.

Of importance nearly equal to the radar picture is the ability to pick up beacons (page 353) which give recognizable landmarks. The set includes a beacon receiver which is tuned by automatic frequency control.

Ship Navigation

Ship navigation is a very well-established art. Nevertheless it is an art into which radar fits perfectly. Microwave ship radar was in production before Pearl Harbor in spite of the fact that the first microwave radar employing a single antenna was not working before March 1941. There is essentially only one hard problem connected with radar for ship navigation, namely the roll, pitch, and yaw of the ship, which necessitates a special design of antenna. Ideally one should have a stabilized antenna for this purpose, and such antennas have been made in quantity. However, for some applications a far simpler solution is to design the "dish" to give a beam which is fanned out by 10 degrees or so, and which accordingly gives adequate illumination of the surface of the sea even when the roll and pitch are considerable. As the refinements in ship radar become of more importance other problems arise. These are notably problems of discrimination, the most important being that of minimum range, which for putting a ship into a dock in thick fog should be a matter of feet. This can be done, but the use of very narrow pulses and wide bandwidths requires that the power consumption of such a radar be considerable.

The use of narrow pulses and narrow beams makes a ship radar a very satisfying instrument. The detailed outline of a harbor can be seen, including the individual piers, and it is even possible to estimate the size and heading of other vessels within a range of a mile or two. Two PPI pictures taken with such a radar are shown in Fig. 11·10.

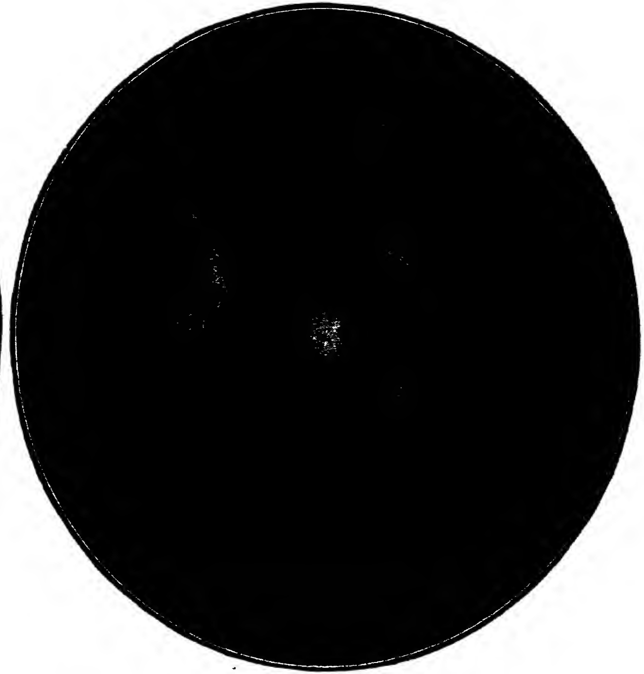
One factor of great importance in ship radar design is the maximum range, which depends on the height of the antenna and the wavelength of the radar. Both control the degree of illumination on the water at any range; the shorter the wavelength the better. Somewhere between 1.3 and 3 centimeters is optimum, the lower value being set at the point at which water vapor absorption becomes serious.

An experimental set built at the Radiation Laboratory and named "Cindy" is as good for illustrative purposes as any. The design characteristics are

Wavelength	1.25 cm (should be slightly longer to avoid water absorption)
Output power	30 kw
Pulse length	⅙ μsec
Antenna size	54 in. by 4 in. (effective), 15 in. (actual)
Receiver sensitivity	15 db down, 10 mc bandwidth
Type of indicator	5 in. PPI
Weight	550 lb
Power consumption	750 watts
Rate of scan	6 rpm

Such a radar could readily be engineered into a reliable set which would need little maintenance. It would not be difficult to incorporate a warning device which would ring a bell when some object was picked up within, say, a mile. Such a device would materially reduce the cost of operation of commercial shipping. The radar installation (not of the type of "Cindy") on the *Queen Elizabeth* saved twice its cost on early trips simply by allowing the ship to proceed at full speed in foggy weather. There is no excuse for collision if one of the vessels has an operating radar, and within a few years it will be proper to say that there is no excuse for an ocean-going vessel not to be equipped with radar. Several

(a)



(b)

FIG. 11·10 Photographs of the PPI of "Cindy," illustrating the high resolution obtained with this radar. (a) Boston Harbor; sweep length about 1 mile. The ship is just north of Fort Point Channel, the bridge across which is clearly seen. The piers of the Charlestown Navy Yard, at the mouths of the Charles and Mystic Rivers, are visible at 1700 yd and 335°. Commonwealth Pier and the Boston Fish Pier, in South Boston, are at 700 yd and 160°. (b) New Haven Harbor; sweep length 2 miles. The ship is in the channel about 1 mile west of Morris Cove, in which the small craft moored at a yacht club are visible. Note the lines of buoys marking the channel, starting at the opening between Ludington Rock and East Breakwaters 1.5 miles south of the ship. The amusement pier at Savin Rock is visible at 1.6 miles and 260°.

installations of radars [12] operating at wavelengths of about 3 centimeters have proved very useful in operations on inland waterways. For example, it has been demonstrated that a tug handling railway barges in New York harbor can continue normal operations under fog conditions which ordinarily bring shipping to a practical standstill.

Air Traffic Surveillance and Control

This very considerable problem requires a radar of good resolution and high performance. The first satisfactory design of such a radar was the MEW (Microwave Early Warning). The official title for this set is the AN/CPS-1.

The design characteristics are

Wavelength	10.3–10.8 cm
Output power	900 kw at 320 pps
Pulse length	1 μsec
Antenna size	Lower beam, 25 ft by 8 ft; upper beam, 25 ft by 5 ft
Antenna gain	Lower beam 10,000; upper beam about 3500, with cosecant squared beam shaping
Receiver sensitivity and bandwidth	12 db down, at 2 mc bandwidth
Type of indicator	Multiple; off-center PPI for control; B-scope for reporting.
Weight	Up to 50 tons depending on complexity of installation
Power consumption	25 kw
Rate of scan	Up to 4 rpm

The appearance of the off-center PPI is shown in Fig. 11·11; this picture was taken during a hurricane in Florida.

One very important requirement for a set of the MEW type is the ability to keep the track of an aircraft once it has been located. This means that the radar has to have an effective range of 170 miles or so in order to give solid response within 100 miles. In addition it is necessary to be able to observe both low and high aircraft. This is achieved by two beams. The coverage given by the two beams is shown in Fig. 11·12. It can be seen that for a

[12] *Railway Age*, **123**, No. 21 (1947); B. B. Talley, *Electronics*, **20**, No. 10, 113 (1947).

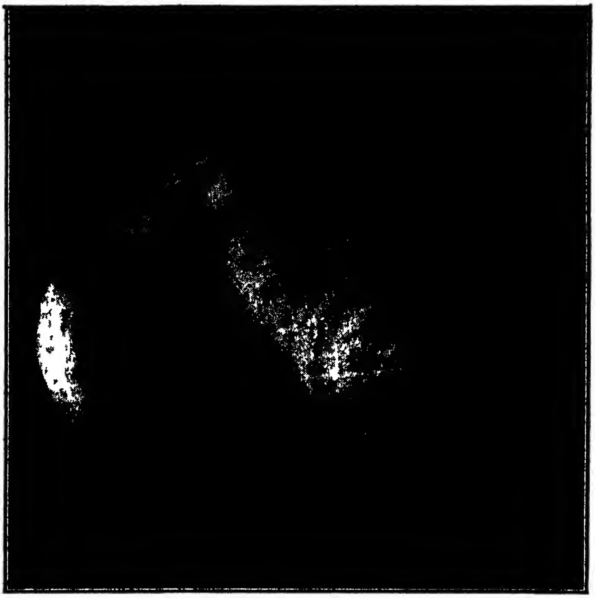


FIG. 11·11 A photograph of the off-center PPI of an MEW radar, taken during a hurricane in Florida. The strong return from the storm clouds and/or rain clearly shows the circular nature of the storm and allows the center of the storm to be accurately located.

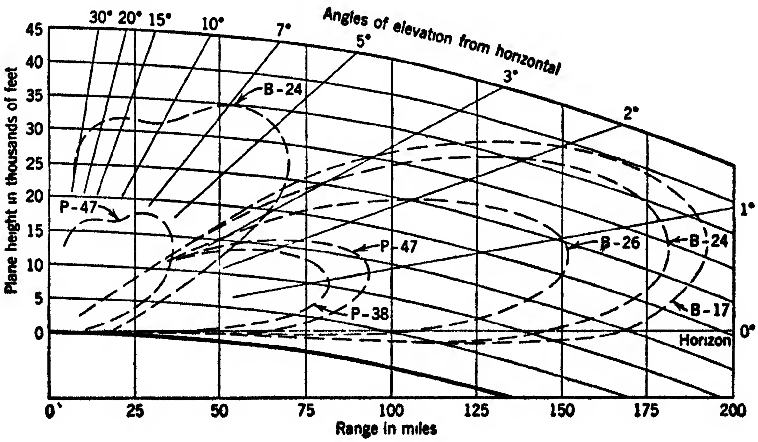


FIG. 11·12 Coverage diagram of the MEW on its two beams for various types of aircraft.

large enough aircraft there is no gap in the coverage, but that for small aircraft there is a gap. Still higher power and better receiver

FIG. 11·13 A photograph showing the superposition of 30 azimuth sweeps on the PPI of an MEW radar located at Tarpon Springs, Florida. The rings are 10-mile range marks. The 30-fold exposure of fixed echoes renders them greatly overexposed, hence the Florida coastline is not clear. The tracks of aircraft, however, can be clearly seen. The sweep rate is 2 rpm, so that the speed of aircraft can be readily estimated. Track A, for instance, covers 10 miles in 6 sweeps and is at 200 mph. The concentrated flying 37 miles south of the radar is activity near an airfield. The large white echoes are local rainstorms, which are very common in Florida in the summer. This picture, taken by members of the MEW group from the M. I. T. Radiation Laboratory, showed for the first time that reasonably gap-free coverage could be obtained by a ground-based radar.

sensitivity are needed. The last remark shows the great importance of high set performance. The output power, 900 kilowatts, represents the limit of the HK-7 magnetron. The HP-10 should

enable this figure to be substantially increased. The receiver sensitivity is capable of some improvement.

A second requirement is the ability to assimilate and use the very great amount of information provided by the radar. To do this, a scheme by which a vertical plotting board is kept filled with tracks by plotters receiving information from five B-scopes was devised. In addition, for control purposes, a number of off-center PPI's were used. The controllers sat at their scopes, with the general situation visible on the vertical plotting board, and took detailed control of any aircraft as requested by the chief controller. The tracks of aircraft can clearly be seen in Fig. 11·13 which is the superposition of 30 rotations of the MEW antenna.

These two simple requirements dominate the design of the set. It is massive, but that cannot be helped if it is to give continuous tracks and be capable of controlling twenty or more separate flights. Its cost is below $100,000 and it can serve a radius of 80 miles. Five sets on the Atlantic Coast would enable aircraft to be controlled all the way from Boston to Washington. There is no doubt that it could be made lighter, yet there is the question whether this is worth doing since the set goes in as a permanent installation.

The MEW needs careful siting since it readily "sees" the ground, including large objects like water towers. It should be placed in rather flat country where the ground horizon is not too far away. The aircraft can then be seen over the ground horizon, while the ground clutter is limited.

Lightweight Height Finder and Bad Terrain Radar

The MEW as just described is excellent for giving the plan location of aircraft but it cannot give height figures. One way of measuring height is to use an antenna with a beam narrow in the vertical direction which is scanned in elevation at a fairly rapid rate. This is done in the "Little Abner" radar, which has an elliptical paraboloid which can be swung up and down through 20 degrees. The echo is shown on a Range Height Indicator (page 209). While performing the purpose of height finding the set can also be used to search for aircraft in regions where heavy echoes from large blocks of ground clutter up the screen of the

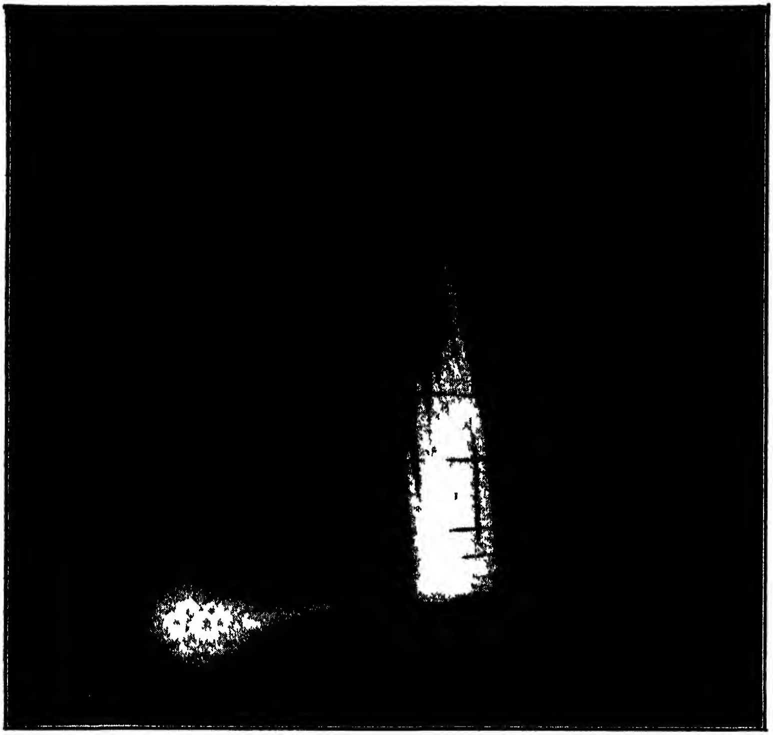


FIG. 11·14 Photograph of the Range Height Indicator of a "Little Abner" radar, showing storm clouds near Iwo Jima extending to an altitude of nearly 28,000 ft.

search set. An aircraft echo is seen to be separated from the land clutter on the RHI and so is detected.

The design characteristics are:

Wavelength	3.2 cm
Output power	80 kw at 800 pps
Pulse length	1 μsec
Antenna size	10 ft by 3 ft cut paraboloid, horn fed
Antenna gain	18,000
Receiver sensitivity	14 db down at 2 mc bandwidth
Type of indicator	RHI
Modulator	Hydrogen thyratron
Weight	2500 lb
Power consumption	3 kw
Rate of scan	Up to 2 per sec (in elevation)

The appearance on the RHI of some storm clouds near Iwo Jima is shown in Fig. 11·14. The set has a range of about 50 miles on a four-motored plane. The narrow beam width of 0.4° enables quite accurate height measurement. Relative heights can be determined with an error of about 300 feet at 30 miles. Absolute heights cannot be given with this accuracy as atmospheric refraction can give angular uncertainties up to half a degree.

Such a set is so simple in principle that all the design effort goes into obtaining range performance on the one hand and ensuring light weight and ruggedness on the other. The use of a slatted antenna is of interest. If the antenna is made with rods or slats parallel to the plane of polarization and closer together than the allowable waveguide width, the radiation will not penetrate between the slats and is reflected very nearly as perfectly as from a sheet of metal. This helps in lessening weight and in reducing wind resistance.

Radar for Guiding Approach to an Airfield

The problems so far described are relatively simple and unspecialized. The problem of bringing an airplane in close enough to an airfield in bad weather to enable it to find the runway is much more demanding. The problem of actual landing is harder still. A radar which comes close to being able to control planes to an actual landing, without requiring any special equipment in the aircraft, is the GCA (Ground Control of Approach). In order to do this, three separate radar antennas are necessary. One of these fans rapidly up and down and provides an accurate picture of the range and height of the plane. A second has a beam which fans back and forth across the runway to give range and azimuth. When the aircraft is within the scan of these two precision radars its course can be accurately plotted and instructions given to the pilot by radio regarding the correct approach.

This simple method of plotting runs into technical difficulties which have forced the actual set to be somewhat complex. In the first place it is necessary to have a nearly continuous record of the position of the aircraft. This means that the scanning has to be done by the precision radars several times a second. In order to avoid serious scanning loss a rapid scan covering a narrow range of angles, supplemented by a servo antenna control, is used. In

this way the operators can slowly move the precision antennas to keep the airplane in the beams.

In addition it is found essential to have a radar for general guidance with much less precision but with greater range and with a 360° search scan. This radar is simple in design except for the antenna which produces a strongly fanned beam to give a coverage up to 4000 feet as close as 4 miles and as far as 30 miles.

The GCA set as used in the war had a rather formidable complement of circuits. It employed four special precision indicators and one ordinary PPI. The precision indicators displayed maximum ranges of 10 miles and 2 miles respectively for the azimuth and elevation radars. The correct line of approach is marked on each of these tubes, and an operator controls a marker which is set mechanically to bisect the aircraft echo on the tube in range and angle. This mechanical tracking is used to feed voltage data to an error indicator which is in front of the controller, who sees in the simplest possible way any faults of piloting and at once radios instructions to the pilot to correct them. The two scope operators have foot pedals with which they can "servo" their antennas to keep the aircraft fully illuminated by the radar beam.

Technically the most interesting feature of the set is the rapid scan process. This is of the delta-A type already described briefly in Chapter 4. The antennas are cylindrical paraboloids fed by dipole arrays in waveguide. The guide wavelength is varied by varying the width of the guide by means of a cam, which is linked mechanically to a variable condenser which in turn transmits the angle information to the rapid scan indicators.

The design characteristics are as follows.

	Search Radar
Wavelength	10.7 cm
Output power	40 kw at 2000 pps
Pulse length	1 μsec
Antenna size	8 ft by 3 ft
Antenna feed	Series of dipoles to give cosecant squared beam
Receiver sensitivity	15 db down at 2 mc bandwidth
Modulator	Hard tube
Type of indicator	PPI
Rate of scan	20 rpm

Precision Radar

Wavelength	3 cm
Output power	15 kw at 2000 pps
Pulse length	0.5 μsec
Beam widths	Azimuth 0.6° by 1.5°; elevation 3° by 0.4°
Receiver sensitivity and bandwidth	16 db down at 2 mc
Type of indicator	RHI type, two for each antenna
Modulator	Hard tube
Rate of scan	Azimuth 20° at 3 per sec; elevation 7° at 6 per sec

This set has had an excellent record, having saved many lives. Some difficulty was met in imparting to pilots the same faith in the laws of physics as the designers had, and for some time after its installation the GCA had to contend with the instinctive distrust of pilots for the ideas of anyone else. This has meant that in many cases its first use has been in very difficult approaches when the pilot has been very scared indeed. That it has succeeded so well under these conditions is a striking tribute to its performance.

The GCA should have great peacetime use. In spite of its elaboration, it is a safety device without equal for aircraft which cannot be specially equipped with landing indicators. With faith in its method of operation and practice in its use before an emergency requires it, a pilot can bring an aircraft in to a landing under GCA control so that vision is needed only for touching down. This covers the vast majority of cases. Where the landing conditions are hopelessly bad, the aircraft should be taken under control by a radar of the MEW type and sent elsewhere. The two sets together could well make air navigation and landing essentially safe in all but the worst weather.

11·4 RADAR BEACONS [13]

A radar beacon, or "racon," is a transmitter which is triggered by signals received from a radar. If the r-f pulse transmitted by a distant radar falls within the pass band of the receiver of the

[13] R. D. Hultgren and L. B. Hallman, *Proc. I.R.E.*, **35**, 716 (1947).

beacon, is of sufficient amplitude, and meets (in some cases) definite specifications with regard to pulse width or some other type of coding, the beacon receiver changes the r-f pulse to a video pulse which triggers the beacon transmitter. The transmitter then radiates either a single r-f pulse, or a coded set of pulses, which will be received by the radar if the beacon transmitter is tuned to the radar receiver frequency.[14]

One of the most important points about beacons is that both the signals from the "interrogating" radar and the reply from the beacon undergo only inverse square law attenuation (under conditions of free space propagation), as compared with the inverse fourth power attenuation which applies to radars. Furthermore, the power radiated by a beacon is in general much larger than that scattered by a radar target. It therefore follows that beacon signals can be detected by radars to considerably greater ranges than ordinary targets. This is true in spite of the facts that beacon receivers are usually considerably less sensitive than radar receivers, and that beacon antennas are in general non-directive in azimuth and only slightly so in elevation, so that they have relatively small gains.

Applications of Radar Beacons

Beacons were originally developed as aids to air navigation. For this purpose, ground-based beacons with relatively high transmitter power (up to 15 kilowatts) are located at appropriate points. The beacons are interrogated by radars in aircraft, the beacon replies being coded so that recognition of the individual beacons is possible. The coding is usually accomplished by means of a set of three or more pulses with characteristic time (or "range") spacing.

[14] In some cases, the radar is equipped with a special receiver tuned to the beacon transmitter frequency, which may then be different from the radar frequency. This arrangement has several advantages which make the added elaboration worth while in certain applications. A ground-based radar can receive beacon signals from aircraft without interference from ground clutter. Beacon transmitters can be set to a specified frequency, and their signals can then be received by radars equipped with special beacon receivers regardless of changes in radar frequencies necessitated by changes of magnetrons or other causes, provided, of course, the beacon receivers have sufficiently wide pass bands to accommodate the radar interrogating signals.

The control of the flight of an airplane by means of an observer stationed at a ground- or ship-based radar and communicating by radio with the pilot is greatly facilitated if the airplane is equipped with a beacon. Not only is the maximum range at which the aircraft can be solidly detected much increased, but also the use of a coded reply makes identification of the aircraft under control much more certain. Furthermore, as mentioned earlier, if the beacon reply frequency differs from the radar frequency and is received by the radar on a separate receiver, there is no difficulty in controlling the aircraft in regions of heavy ground or sea clutter. Beacon-aided control of aircraft was much used in the war for such purposes as air reconnaisance and air support of troop operations, MEW and 584 radars serving as the control radars. Figure 6·6 illustrates the appearance of a coded beacon reply on a PPI when either the beacon frequency is the same as the radar frequency, or the video signals from a separate off-frequency beacon receiver at the radar are mixed before display with the radar video signals.

Lightweight beacons have been developed which are useful in temporarily marking certain locations. Such a beacon can be dropped by parachute from an airplane to aid in guiding other planes to the same point, as in paratroop operations. A beacon installed at an accurately known position relative to a nearby enemy installation can be used in radar control of gunfire. It is evident that such marker beacons can find peacetime applications in radar-aided mapping and surveying.

Racons proved to be very useful in certain types of bombing operations. For example, the so-called "oboe" system, which was employed with great success in the strategic bombing of the Ruhr district, employed two ground-based interrogating radars and a beacon-equipped aircraft, which could serve as the pathfinder for several bombers. One of the radars, the "cat" station, held the plane, by means of radioed instructions, on a circular course which would carry it over the bomb-release point. The second ("mouse") station was located at some distance from the cat station, and gave the bomb-release time to the plane when the latter was at the proper distance from the station. The importance of this system was that the precision of the operation depended, so far as the radars were concerned, only on their range accuracies, so that no high degree of directionality in the radars was required.

IFF

In many military operations it is of the greatest importance to be able to distinguish friendly aircraft and ships from enemy ones. The IFF (Information Friend or Foe) equipment used for this purpose resembles in some respects radar and beacon equipment, and is therefore briefly mentioned here. Because of the wide diversity of radar frequencies employed, ranging from 100 to 10,000 megacycles, it was obviously impossible to accomplish universal identification by means of racons. IFF gear, operating at a few hundred megacycles, was composed of an "interrogator-responsor" located at a radar, and a "transpondor" located at the target to be identified. The former unit served to trigger the latter in the same way a radar triggers a beacon, and to detect the coded response transmitted by the transpondor. The detected IFF signals were mixed with the radar signals and displayed on the radar indicators. The IFF signals served only as aids to identification, since the range and angle selectivities of the radar were in general much greater than those of the IFF equipment.

Loran [15]

Loran (*Lo*ng *Ra*nge *N*avigation) is a system of navigation, which, although it bears little resemblance to radar, is of such great importance that mention should be made of it here. The unit of Loran operation is a pair of stations located several hundred miles apart and transmitting pulsed signals. The pulses from the so-called slave station are synchronized with the pulses from the master station. A navigator measures the time difference between pulses received from the two stations, and thus determines the difference between the distances to the stations. The locus of points having a constant value for this difference is a hyperbola. A similar determination using the pulses transmitted by a second pair of Loran transmitters gives a second hyperbola, the intersection of which with the first one defines the navigator's position. The various pairs of stations are distinguished by slightly different pulse repetition frequencies, in the neighborhood of 25 pulses per second. The navigator adjusts the sweep fre-

[15] B. W. Sitterly, Radiation Laboratory Report 499, May 1944; J. A. Pierce, *Proc. I.R.E.*, **34**, 216 (1946).

quency of his indicator to cause the pulses from the desired pair of stations to form a stationary pattern on the indicator.

Loran transmitters operate on a frequency of approximately 2 megacycles and radiate a peak power of the order of 100 kilowatts. Loran signals are propagated by both ground waves and sky waves (reflections from the ionosphere), and, under certain restrictions, signals propagated in either way are useful. Standard Loran, depending on ground waves, gives the greater accuracy (better than a mile) but can be used only to ranges of several hundred miles. SS (*S*ky Wave *S*ynchronized) Loran, which employs signals propagated by sky waves, can be used at distances up to 1500 miles; it is generally useful only at night and is unreliable at ranges less than about 300 miles from a Loran station.

Loran is mainly of use for maritime navigation by both aircraft and ships. It will undoubtedly establish for itself an important permanent position in the science of navigation.

11·5 PRECISE ANGLE DETERMINATION; AUTOMATIC TRACKING

The problem of making highly precise *range* measurements by means of radar presents no special difficulties. The use of very short pulses and precision, crystal-controlled ranging circuits has given accuracies of the order of 20 yards. An analogous approach to precise angle measurements, which consists in forming an extremely narrow radar beam and making an accurate measurement of antenna position when the radar is "on target," is not practicable because of the unwieldy dimensions of an antenna giving a sufficiently narrow beam, and, of greater importance, the practical impossibility of finding the desired target with a very narrow beam. Experience has shown that there is some difficulty in picking up an aircraft target with a beam of circular cross section even when the beam width is as much as 4°. In fact, radars having azimuth resolution down to 1° or less have been successful only because they have employed fan-shaped beams which essentially require scanning in only one angular dimension instead of two.

A much more powerful method of gaining angular accuracy depends on echo strength comparisons. As a simple illustration, we may consider a radar with an antenna pattern which is fanned out in elevation and a few degrees wide in azimuth. The antenna

can be equipped with two feeds which give slightly different beam directions relative to a horizonal line perpendicular to the antenna, and an r-f switch can be arranged to send alternate transmitter pulses to the two feeds. If a target is located so that it is better illuminated by one beam than by the other, alternate received echoes will be larger. An A-scope presentation can then be employed to adjust the antenna position until all the received pulses are of equal amplitude. In this way angles can be measured to considerably less than the beam width. It is, of course, necessary that the two beams overlap so that the target can be detected in each simultaneously. The angle of beam separation is usually about equal to the beam half-power width (for two-way transmission), so that "crossover" occurs at an echo amplitude of about seven-tenths maximum.

A much more refined method of echo strength comparison is employed in some precision radars. For example, the SCR-584,[16] which operates at a wavelength of about 10 centimeters, has an antenna consisting of a parabolic reflector illuminated by a rotating off-centered dipole feed. Rotation of the feed causes the beam, which is roughly 3° wide (two-way transmission), to describe a cone. As a result of this conical scanning, the signals returned from a target which is off the axis of the paraboloid will be more or less amplitude-modulated at the frequency of rotation of the feed, in this case about 30 cycles. If crossover occurs at about one-half maximum power, the echo from a target one-half degree off the paraboloid axis will vary between the approximate limits 0.2 to 0.9 of the amplitude which would be observed with a stationary beam trained on the target. The 30-cycle modulation can be detected, and the antenna pointed to remove it. In order to supply information as to which direction the antenna should be turned to remove the modulation, the *phase* of the modulation signal is compared with the phase of a 30-cycle reference signal from a generator rotating at the same speed as the antenna feed. One can see that a not too formidable elaboration of equipment is involved in supplying the modulation error signal, after suitable phase comparisons, to azimuth and elevation servomechanisms to cause the antenna to position itself automatically to remove the echo modulation. This automatic following is actually accomplished, with a precision of something like 0.05°, in the 584. The

[16] *Electronics*, **18**, 104, Nov. 1945; **18**, 104, Dec. 1945; **19**, 110, Feb. 1946.

angle data thus derived can be fed to a computer for training an anti-aircraft gun, or can be used directly for training a searchlight, or supplied along with range data, to an automatic plotting board for accurately recording the flight of an airplane. Incidentally, a later modification of the 584 includes circuits for accomplishing automatic range measurements, so that once a target is brought within the pull-in region of the radar, continuous and accurate range, azimuth and elevation data are fed to gun computers or other devices, without any operator attention! It is truly impressive to see such a radar in operation.

11·6 MOVING TARGET INDICATION

A very serious problem of radar is the elimination of the strong echoes from ground objects without removal of the rather weak echoes from aircraft. A method for doing this was worked out by Grayson in England, and McConnell and Emslie at the Radiation Laboratory. The principle is simple; the practical application is difficult.

The process involves the fact that the aircraft is moving and the ground is not. It would seem that this could easily be applied as follows. The series of ground echoes in one repetition interval could be stored in some way and then reversed in sign and used to cancel the echoes in the next interval. Only the aircraft echo which is changed in location would not be wiped out. This simple idea does in fact form part of the process of moving-target indication. However, the time allowed for motion is very short, and the pulse length ordinarily allows only a discrimination of the order of a hundred yards or so. The aircraft does not move a hundred yards in a thousandth of a second. In consequence the aircraft echo is essentially at rest and so is also cancelled.

This difficulty is overcome by using transmitted pulses the phase of which is known with respect to a steady c-w oscillator. The echo from ground or aircraft can be made to "interfere" with this steady oscillator and, as a result of the phase relationship between echo and oscillator, will give maxima or minima. The phase relationship is determined by the distance of travel of the radio waves and is fixed for ground objects. On the other hand it is changing for a moving object like an aircraft, so that the pulsed part of the

resultant of the echo and the steady oscillation fluctuates very rapidly. If now we apply the process of storage, reversal, and cancellation to the video signals derived from this resultant, the fixed objects are cancelled, and moving objects are not, in general, cancelled. This means that there is an almost complete removal of ground echoes while aircraft echoes are only removed in the extremely unlikely event that the aircraft travels at a speed such that the phase relationships give a constant mixed echo.

The method of storing the echoes in one repetition interval involves the use of some physical process which travels much more slowly than light and yet quite fast. The sweep of an electron beam across a tube is one such process, and this has been tried. Another is the travel of sound in water or some liquid which is not too temperature-sensitive. In this method, the video resulting from the echoes and the c-w oscillator, known as "coherent video," is used to modulate a 10- or 15-megacycle carrier, which then drives a crystal and generates supersonic waves in a column of liquid of precisely the right length to produce a delay of one repetition interval. After detection, the delayed video is reversed in sign and mixed with undelayed video, with the result that cancellation occurs for all but aircraft and moving objects.

REFERENCE

A compact summary of the microwave radar art is given by E. G. Schneider, *Proc. I.R.E.*, **34**, 528 (1946).

CHAPTER 12

MICROWAVE COMMUNICATIONS

The rapid development of microwave techniques during the war has opened up a whole new region of the electromagnetic spectrum to exploitation for purposes of communications. The important properties of microwaves and microwave equipment make it certain that we can expect widespread developments in this direction in the next few years, developments which will undoubtedly produce some startling applications. The most obvious advantages offered by microwaves are bandwidths of hundreds or even thousands of megacycles, highly directional antennas, and the wide frequency regions available. Large bandwidths bring the possibility of transmitting information at a very high rate over a single radio link. Thus, a very large number of channels of speech or music, which may be classed as narrow band intelligence, may be handled simultaneously by a pair of communicating stations, or one channel (or possibly a few channels) of broadband intelligence such as that involved in color television. Narrow radio beams result in lowered power requirements, improved secrecy of transmission when this is desired, and decreased interference from other transmitters. With a very wide frequency region available it will be possible materially to relieve the congestion of the "ether" in spite of adding many new channels. These considerations lead to the expectation that the most important industrial application of microwaves will lie in the communications field.

The development of microwave communications is at present receiving intensive industrial attention, but under conditions which

make it impossible for the present authors to claim any familiarity with the rapidly changing state of the development. Our discussion in this chapter will therefore necessarily be limited to pointing out some of the general possibilities and limitations as indicated by the basic physics of the subject, and to a brief description of some of the wartime applications of microwave communications.

12·1 PROPAGATION

The limitations imposed by propagation in the communications application of microwaves are quite different from those in the radar application, discussed in Chapter 11, chiefly because of two factors. In the first place, communications systems involve only one-way transmission; therefore the inverse fourth power dependence of signal strength on distance which applies to radar becomes an inverse square law. This considerable advantage in favor of the communications case is counterbalanced by a second factor, that of reliability of reception. Radar echoes only slightly larger than receiver noise are perfectly usable, whereas signals conveying audio or picture information must in general be much larger than noise, particularly in the entertainment field. Furthermore, in the radar case, especially with long wave radars, it is expected that under some circumstances a given radar signal will "fade" to disappearance at more or less frequent intervals; such fading cannot be tolerated in many communications applications. In order to be certain to receive a fully usable signal even during periods of severe fading it is necessary to design a communications link so that the received signal is of the order of 50 decibels above noise under normal conditions.

The limitation of high frequency propagation to line-of-sight paths requires the use of high antennas to achieve useful ranges. According to equation 11·15, if the transmitting and receiving antenna heights are equal, in order to have a line-of-sight path of r miles over a smooth spherical earth the antenna heights necessary are

$$h_t = \frac{r^2}{8} \tag{12·1}$$

Experience has shown that the line-of-sight requirement is not sufficiently conservative for reliable communications links. Han-

sell[1] points out that interference effects may be troublesome, but that these can be avoided, in the case of a perfectly reflecting spherical earth, by choosing antenna heights so that the path for the reflected beam is no less than a sixth of a wavelength longer than the path for the direct beam. If a path phase shift of 60° is subtracted from the reflection phase shift of 180°, the resultant of the direct and reflected beams will have the same intensity as the direct beam alone. If the antenna heights are further increased the intensity of the resultant will exceed this value over a wide range of heights. Table 12·1 gives approximate antenna heights required to satisfy the simple line-of-sight criterion and Hansell's criterion for various distances at 500 and 8000 megacycles.

TABLE 12·1 ANTENNA HEIGHTS FOR LINE-OF-SIGHT AND HANSELL'S CRITERIA

Frequency (megacycles)	Transmission Distance (miles)	Antenna Heights	
		Line-of-sight, h_t (feet)	According to Hansell's criterion, H_t (feet)
500	10	13	110
500	25	80	230
500	50	330	550
8000	10	13	30
8000	25	80	120
8000	50	330	390

Even though there is very little reflection from ordinary land surfaces at frequencies above about 2000 megacycles, for angles near grazing incidence, it has been found empirically[2] th micro-wave transmission becomes much more subject to severe fading when there is an obstruction in the path causing grazing or almost grazing incidence. Durkee[2] suggests that over rough terrain a

[1] C. W. Hansell, *Proc. I.R.E.*, **33**, 156 (1945); L. E. Thompson, *ibid.*, **34**, 936 (1946).

[2] A. L. Durkee, Bell Telephone Laboratories Report MM-44-160-190, Aug. 1944.

practical criterion for essentially free space propagation is that the first Fresnel zone of the beam should be clear of all obstructions. The first Fresnel zone at a distance r from a source of radiation is the circular area on a plane perpendicular to the line of propagation bounded by the intersection of the plane with a sphere of radius $r + (\lambda/2)$ with its center at the source. It is

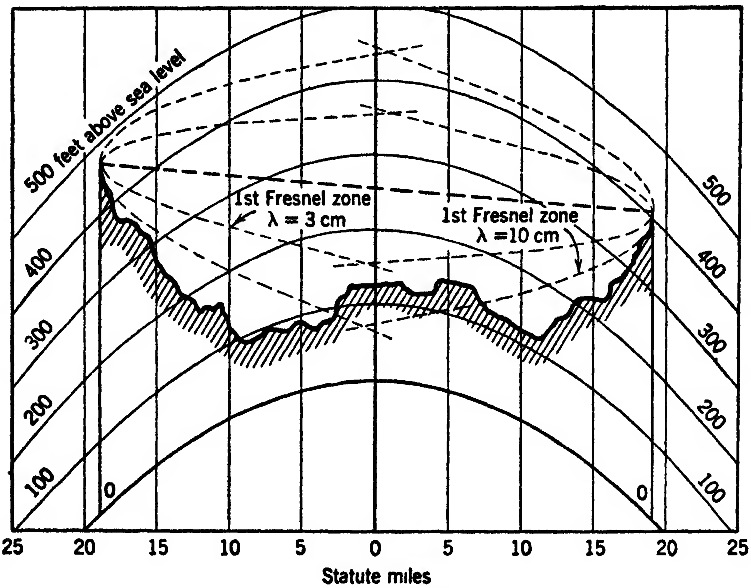


FIG. 12·1 A hypothetical transmission profile. A clear line-of-sight path is available; according to Durkee's criterion the path should be fully satisfactory at 3 cm but not at 10 cm.

evident, from the fact that one can interchange transmitter and receiver without affecting the propagation in any way, that r should be taken as the shorter of the two distances from the obstruction to the ends of the transmission path. It is easily shown that the diameter d, in feet, of the first Fresnel zone at a distance r miles is given by $d = 26\sqrt{r\lambda}$, λ being expressed in centimeters. Since in all practical cases this radius is small compared to the beam width, no consideration has to be given to the shape of the beam. Figure 12·1 shows a schematic transmission profile for two stations mounted on the tops of small hills about 40 miles apart. It is seen that this path should be satisfactory for $\lambda = 3$

centimeters but unsatisfactory or marginal at $\lambda = 10$ centimeters, according to Durkee's criterion.

The formation of trapping layers (page 331) and other atmospheric conditions which affect refraction at radiofrequencies may well lead to very severe fading at some times of the year and in certain regions of the earth. The interesting observations reported by Durkee [2] indicate that, at least in the northeastern part of the United States, such effects are apt to be somewhat less serious at microwave frequencies if path lengths do not exceed about 30 miles.

Severe fading can result from interference effects in transmissions between an aircraft and a station at the earth's surface, particularly over water. This difficulty can be alleviated in the case of transmission from an aircraft to a ship by the use of a so-called diversity receiver,[3] since the interference lobes at the receiving end are in this case closely spaced. A diversity receiver consists, in one form, of several receivers fed by antennas mounted vertically one above the other; the receiver outputs are mixed and, to avoid unduly high receiver noise, an automatic gain control is employed which controls the gain of all the receivers but is itself only controlled by the receiver which is receiving the strongest signal.

Absorption of very high frequency radiation by the atmosphere was discussed in Chapter 11. This discussion applies equally well in the present connection. It appears that absorption effects may become important at frequencies much above 10,000 megacycles, but that at frequencies lower than this trouble should arise only in periods of rainfall of cloudburst proportions.

12·2 POWER REQUIREMENTS

Very high frequencies have an important advantage where point-to-point communication is involved in that a large fraction of the power required at lower frequencies can be replaced by considerably less expensive antenna gain. A limitation on usable antenna gain is imposed by the difficulties in properly lining up two antennas having very narrow beams and maintaining that alignment in heavy weather.

[2] A. L. Durkee, Bell Telephone Laboratories Report MM-44-160-190, Aug. 1944.

[3] F. E. Terman, *Radio Engineers' Handbook*, McGraw-Hill Book Co., 1943, p. 660.

Suppose a radio link is to operate at 10,000 megacycles, with parabolic antennas 4 feet in diameter at both ends of the link. At this frequency there is no difficulty in providing a receiver having a noise figure (page 245) of 12 decibels. If the bandwidth is to be 10 megacycles, a received signal 50 decibels above noise has a power P_r of $10^5 \times 6.4 \times 10^{-13}$ watt. The relation between transmitted power and received power is seen, by an obvious modification of the treatment given in Chapter 11, to be

$$10 \log \frac{P_r}{P_t} = db_p{}^0 + 10 \log \frac{G^2\lambda^2}{16\pi^2}$$

where P_t is the transmitted power and G is the antenna gain. In this case $G \approx 9400$, if we take the antenna efficiency factor to be 0.6. Therefore

$$10 \log \frac{P_r}{P_t} = db_p{}^0 + 27.1$$

For a range of 25 miles, $db_p{}^0$ according to Fig. 11·3 is -92, so that

$$10 \log P_t = 92 - 27.1 - 71.9 = -7$$

or

$$P_t = 200 \text{ milliwatts}$$

This amount of power is nearly within the capabilities of a small tube such as the 419 klystron (page 62). We thus see that surprisingly small amounts of power are sufficient for point-to-point radio links operating at very high frequencies. This advantage of microwave carriers will obviously disappear in situations where directional beams cannot be used, as in the broadcast field.

Microwave communications networks have to include numerous relay stations, not only to minimize fading but also to surmount geographical obstacles. Ideally, a relay station consists of a receiving antenna followed by an r-f amplifier feeding into a transmitting antenna. Microwave r-f amplifiers available at present are characterized by rather high noise levels, so that relay stations employing them have to be situated close together to maintain satisfactorily high signal-to-noise ratios. It is possible that tubes such as the "traveling wave" tube[4] will improve this situation.

[4] J. R. Pierce and L. M. Field, *Proc. I.R.E.*, **35**, 108 (1947); R. Kompfner, *ibid.*, **35**, 124 (1947).

An alternative scheme, which involves considerable added equipment, is to demodulate the incoming signal at a relay station, amplify the modulation signal, and use it to modulate a transmitter operating at a different frequency. Microwave communication links involving as many as seven relay stations,[5] are now in operation.

12·3 FREQUENCY STABILIZATION

Microwave oscillators are apt to drift in frequency as a result of temperature or voltage changes to a large enough extent to require some form of automatic frequency control (AFC). The radar AFC scheme described on page 262 may be applied here, after suitable modifications. The gain of the i-f amplifier must be increased to allow for the lower signal strength, since here the control must be based on the received signal, even though this introduces the chance of locking on an interfering signal instead of the desired one. In the case of a frequency-modulated system, the need for a separate AFC i-f amplifier is obviated by the use of the discriminator output for control purposes. After d-c amplification and low pass filtering to avoid degeneration of signal frequencies, the discriminator output is fed to the local oscillator in proper phase to correct for frequency shifts.

The r-f discriminator developed by Pound (page 267) can be employed to give very precise frequency control of both the transmitter and local oscillator.[6] As in the case of more conventional AFC systems, modulation frequencies must be attenuated in the feedback loop.

12·4 MULTICHANNEL COMMUNICATION

Microwave transmitters and receivers at present available can be adapted to handling modulation signals covering a band of the order of 20 megacycles. This gives the possibility of transmitting a very large number of audio channels or several video channels over a single r-f carrier by a variety of methods.

[5] *Electronics*, **21**, 114 (Jan. 1948).

[6] L. M. Hollingsworth, H. Logemann, A. W. Lawson, and J. M. Sturtevant, Radiation Laboratory Report 977, Jan. 1946.

Several methods of multichannel communication are based on time discrimination methods. The pulse time modulation system, in which the time interval between a reference trigger pulse and a movable pulse is varied in accordance with the intelligence to be transmitted, can handle several audio channels if several different movable pulses are used in each repetition interval. In transmitting intelligence by pulses, whether by pulse amplitude or pulse time modulation, the pulse repetition frequency should be higher than twice the highest modulation frequency to be transmitted in order that the lowest combination frequency of the modulation and pulse frequencies can be removed from the signals by a low pass filter. If pulses of the order of a microsecond wide and delay circuits of ordinary precision are employed so that the r-f bandwidth required is only one or two megacycles, the restriction to high repetition frequencies limits to a relatively small number the audio channels which can be accommodated. If, however, very narrow pulses and precision delay circuits are used, with correspondingly large r-f bandwidth, the number of signal channels may be much increased.

Mention should be made here of the newly developed pulse code modulation,[7] which is applied to multichannel communication, and which has the important property of permitting practically noise-free transmission.

An ingenious multichannel communication scheme may be based on the use of storage tubes (page 186). If a train of audio signals covering a 5-kilocycle band and lasting t seconds is stored in such a tube and then removed in $t/1000$ second, it is converted into a 5-megacycle band of signals. In principle, several hundred such audio bands could be compressed into a single 5-megacycle video band for transmission and then expanded back to audio signals at the receiving station. Switching transients would be introduced into the audio signals, but these could be made short enough not to interfere appreciably with intelligibility.

In certain specialized applications, such as radar relay (Section 12·5), it may happen that each of several channels of video or synchronization intelligence needs to be used only during a certain portion of each repetition interval. In such cases simple elec-

[7] W. M. Goodall, *Bell System Tech. J.*, **26**, 395 (1947); L. A. Meacham and E. Peterson, *ibid.*, **27**, 1 (1948).

tronic switching of a single transmitter between the various signal channels may be used.

In connection with multichannel transmission it may be noted that a single microwave antenna of cylindrical symmetry can be used with two transmitters or two receivers by having the polarizations of the two r-f beams mutually perpendicular.[8] One cannot rely on crossed polarization to give more than 15 or 20 decibels decoupling between the two r-f channels, so that it will be necessary in general to use two different r-f carrier frequencies in order to obtain further discrimination by receiver selectivity. Actually, the chief purpose of the crossed polarization is to remove the 3-decibel losses which would otherwise result from having an extra receiver (or transmitter) present to absorb half the energy of a received (or transmitted) signal.

12·5 MISCELLANEOUS APPLICATIONS

In this section we describe briefly a few examples of microwave communications equipment developed for various applications during the war.

Ultraportable Telephone Radio Link

The U. S. Signal Corps developed a microwave transmitter-receiver for audio communications which is unique and interesting in several respects. This equipment was designed primarily to serve as a radio link in a wire telephone system to afford easy bridging of narrow bodies of water or difficultly accessible valleys. The transmitter consists of a single modulation amplifier tube and an amplitude-modulated lighthouse tube (Chapter 3) operating in the low frequency region of the microwave spectrum. The receiver likewise consists of only two tubes, a lighthouse super-regenerative detector and a stage of audio amplification. A small parabolic antenna is used for both transmitting and receiving. The equipment may be battery-operated and is easily portable. A free space range of 5 to 10 miles is reported.

[8] A. W. Lawson in Radiation Laboratory Report 977, Jan. 1946.

Pound's Cross Band System

Pound has developed an interesting system for communication at about 10,000 megacycles, based on his r-f discriminator type of frequency stabilization (page 267). In this system, the transmitters are frequency-modulated, the unmodulated frequency being controlled by an r-f discriminator. The transmitters are separated in frequency by the receiver intermediate frequency, so that each receiver employs as its local oscillator the transmitter at the same station. This system, because of its very high carrier frequency, is ideally adapted to cases where highly directional beams are required; correspondingly, good ranges can be obtained with low power tubes such as the 419 klystron. Wideband signals, such as video signals, can be accommodated.

Radar Relay

It was pointed out in Chapter 11 that radars such as the MEW can collect a tremendous amount of information, a fact which led during the war to the use of numerous indicators with each MEW so that many operators could simultaneously interpret specified portions of this information. The information collected by a radar may also be of importance, for example, to air traffic control officers at some remote location. Because of the rapid motion of modern aircraft, it is evident that radar information concerning them becomes rapidly less significant with time; in many situations, data a minute or so old are practically useless. This consideration shows that telephonic transmission of radar data in general detracts greatly from their usefulness. A much better solution is to give to operators at remote stations exact duplicates of the indicator presentations available at the original radar, and with a negligible time delay. This is the solution aimed at by radar relay.

An obvious method of relaying radar information by radio is to televise an indicator at the radar; this scheme has been proposed for use in air traffic control systems. Although this device is entirely satisfactory for some applications, it is inadequate for others. Numerous specialized types of radar indicators have been developed, some of which are described in Chapter 6. For the fullest utilization of the data from a radar such as the MEW, it is impor-

tant that these various types of presentation be available simultaneously even at remote stations. The television scheme obviously allows only one type of presentation per television channel. A more generally useful scheme is to relay the "raw" radar information, consisting of one or more channels of video signals and data for synchronizing indicator sweeps with the radar modulator pulse and the motion of the radar antenna.

There can be no doubt that radar relay will play a very important role in whatever system, or combination of systems, is eventually adopted to give a safe and efficient solution to the increasingly complex problem of the control of civilian air traffic.

CHAPTER 13

MICROWAVES IN PHYSICAL RESEARCH

In Chapter 11 we have described the use of microwave methods in the construction of radar. In Chapter 12 the possibilities of microwave communications have been discussed. These are not the whole story, although much development has gone into them, and in this chapter we describe four research fields in which there is considerable current activity. More will doubtless appear.

These four lines of research bear very little relation to one another, except in the common methods of microwave manipulation. This is what one would expect in research in which microwaves are used as a tool to gain an end. The grouping of these in one chapter is therefore largely useful to the reader who seeks to see what can be done with microwaves and who would like to use the technique for his own ends.

The four topics are: microwave absorption; radiation measurement by microwave technique; microwave accelerators for nuclear physics; microwaves and superconduction.

13·1 MICROWAVE ABSORPTION

The first discovery of microwave absorption was made by Cleeton and Williams in 1934. Their methods differ considerably from those used today, but they overcame considerable experimental difficulties to secure results which were a clear ten years ahead of the times. Their work will be described more completely later.

At present we can consider why a microwave absorption spectrum is of interest to anyone other than those engaged in filling out tables of numbers.

The physical interpretation of molecules is becoming steadily more and more important. Great advances along these lines have been made in which almost every branch of physics from theoretical quantum mechanics to x-ray and electron diffraction has been used. The aim of this interpretation goes beyond the description of plausible models: it is toward a complete account in which all kinds of separate information are interconnected to give the final description. Thus one can, from measurement of a gaseous dielectric constant, infer a dipole moment, which in turn must fit in with the molecular sizes determined from electron diffraction and also have a bearing on the temperature coefficient of viscosity. The addition of one more source of physical information about molecules is therefore of value even before the available results have been fully examined. This is one reason for the interest in microwave absorption.

A second reason is that microwave absorption enables information to be gained on collision processes, which in turn involve the forces between molecules. This information is obtained from the strong pressure effects observed which are much more pronounced in microwave spectra than in optical spectra, and form, indeed, the most striking feature of absorption measurements in this region of the spectrum.

A third region of interest is the study of the effect of electric and magnetic fields on microwave spectra. It has been shown that this can lead to the determination of nuclear spin among other things.

Many other possibilities exist. Little work has been done on absorption by liquids and more complex molecules. This field may well prove to give figures which are of great use in analysis of structure or, in reverse as is now done for infrared spectra, for physicochemical analysis.

It is of some interest to inquire into the manner in which a molecule can suffer a transition between two energy levels so close together that a microwave absorption results. The energy equivalent to one quantum at 30,000 megacycles or 1-centimeter wavelength is given by $h\nu = 6.6 \times 10^{-27} \times 3 \times 10^{10}$ or 1.98×10^{-16} erg. (Here h is Planck's constant and ν is the frequency.) This is

1.24×10^{-4} electron volt per molecule, or 2.86 calories per mole. This is very much less than the energy change in the mildest chemical reaction between ordinary molecules. It is also far less than the usual energy separation between levels due to electronic transitions or to vibrations in ordinary molecules. On the other hand the energy separation between rotational levels is not so large. For an ordinary linear molecule like OCS (carbonyl sulfide) the moment of inertia perpendicular to the axis of the molecule is about 10^{-38} gram cm^2, and so the energy levels are given by

$$\frac{h^2 J(J+1)}{8\pi^2 \times 10^{-38}} \text{ erg}$$

where J can have the values 0, 1, 2, 3, $\cdots$, etc. The separation between the first two levels is therefore 0.55×10^{-16} erg, which is less than the above figure, corresponding to a still longer wavelength. The separation of rotational levels is therefore one which is of the right order, and it may be expected that in one form or another microwave spectra will be concerned with rotational levels. Measurements on rotational levels yield information about the moments of inertia of a molecule which in turn is a clue to the correct molecular structure. The rotational lines are very sharp, and the moments of inertia so deduced are very accurate.

In addition, many processes cause energy levels which are normally multiple, but have exactly the same energy for each term (so-called degenerate levels) to split into closely separated energy levels. This occurs in the Zeeman effect or the Stark effect. The transition between these closely separated energy levels can well be in the microwave region. Moreover, since the level separation is to some extent under control by the magnetic or electric field, this type of absorption process is attractive to study. The element of chance in finding the correct frequency for absorption is lessened. The precise observation of the way in which the multiplets develop gives valuable information about the precise nature of the molecular energy level concerned, and so is of great interest.

Moreover, it has recently been proved possible to observe the "hyperfine structure" due to the coupling of the electric and magnetic fields of the nucleus with the molecular structure. Such observations tell us the nuclear spin and the nuclear magnetic or electric moment.

Experimental Method

Although, in principle, we should outline the various avenues for research in microwave spectra from the theoretical point of view, this is not possible at the present time. The reality is that the limitations on research imposed by experiment are so great that one is more inclined to seek a problem to fit the equipment than to design equipment to solve a problem. This being so we can here describe some of the present methods of working on microwave spectra and then indicate the theoretical interpretation which has been placed on the results. Two general methods are in use: attenuation measurements in waveguide; and the change in Q of a large cavity.

Attenuation Measurements in Waveguide

A well-constructed metal waveguide has an attenuation on its own account of somewhere near 1 decibel per meter at 30,000 megacycles. This was discussed in Chapter 2. Moreover it is quite constant. Therefore if some method of observing the power transmitted by a waveguide is devised, the change due to the introduction of an absorbing gas will be measurable even if the absorption is small. This technique was first carefully applied at the Radiation Laboratory by Beringer.[1] The element studied was oxygen, and the absorption path was a rectangular silver waveguide of length 6.19 meters, 0.18 inches wide by 0.086 inches deep. The process of absorption is as follows. The electromagnetic field of the radiation acts on the electrons in the molecule. This force is superposed on the force binding the electrons in position in the molecule, and while it is present the molecule is in an abnormal condition. The electrons, and indeed the molecule as a whole, can exist in certain definite conditions, characterized by definite energy values. If the radiation is of high enough frequency so that the value of $h\nu$ is sufficiently large, the abnormal condition of the molecule can result in the transition of the molecule to a state of higher energy. This energy is subsequently re-emitted in one form or another. Such transitions are exceedingly improbable unless the frequency of the radiation is somewhere near the resonant frequency. In any event the transition does not occur in

[1] R. Beringer, *Phys. Rev.*, 70, 53 (1946).

a regular manner; it is a matter of pure chance, although the odds may vary considerably. One cannot say which molecule will absorb the radiation, but only that on the average some will. The rate of average absorption may vary greatly but the random nature is still there. These random processes are conventionally described in terms of a *cross section per molecule.* The cross section is a fictional quantity which represents the effective target area presented by each molecule. For a "probable" process (i.e., one with favorable odds) the "cross section" is large; for an "improbable" process it is small.

If we represent this effective target area per molecule by S and there are N molecules per unit volume, we can express the area of target presented to a beam of radiation of area A passing through a thickness dx as S times the number of molecules present, or $SNA\,dx$. Now we can describe the probability of absorption in two ways: the first is the ratio of area of target presented to the area of the beam, which is $SNA\,dx/A$ or $SN\,dx$; the second is the ratio of the change in beam intensity to the original beam intensity or $-dI/I$. Equating these gives

$$-\frac{dI}{I} = SN\,dx$$

and if we use I as the intensity of the radiation at a distance x, and I_0 as the initial intensity we get

$$\frac{I}{I_0} = e^{-SNx} \tag{13·1}$$

The quantity SN is usually written as α and is called the absorption coefficient. The experiment is therefore concerned with the measurement of α, from which S can be derived.

The block diagram of the equipment used by Beringer is shown in Fig. 13·1. The microwave power from a 1-centimeter klystron is fed into a crystal frequency multiplier in which the frequency is doubled and from which the ½ centimeter power is fed down the waveguide. The power transmitted is detected by a crystal, and the rectified voltage is amplified. In order to make possible the measurement of the relatively small change in transmitted power when gas is introduced, the unattenuated voltage due to the evacuated tube is balanced out, and the change from balance is measured when the gas is introduced.

The gas introduces effects due to its dielectric constant which are not due to absorption. If the standing wave ratio is low the total effects are less than 0.1 per cent. One effect is due to the change in the field and current distribution in the guide. This changes the way in which the skin effect currents are flowing and

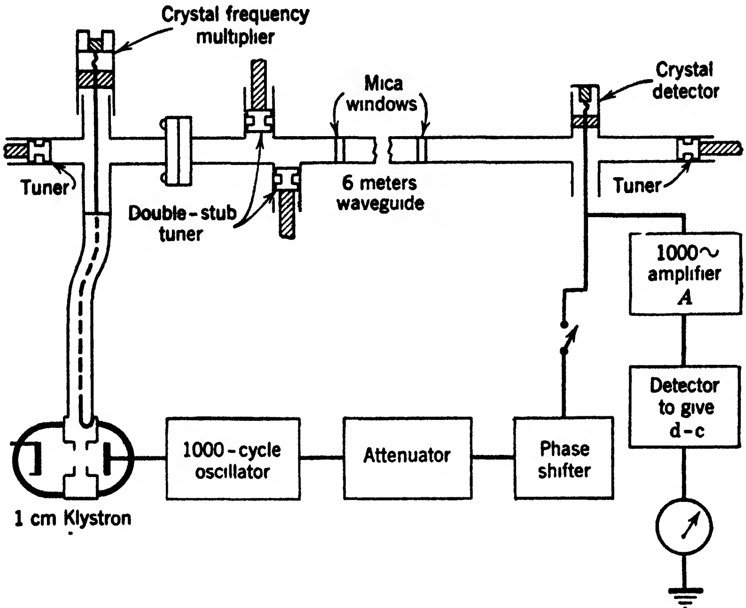


FIG. 13·1 Schematic diagram of Beringer's original apparatus for measurement of microwave absorption. Radiation at ½ cm, which is the first harmonic of a 1-cm oscillator, is sent through a section of waveguide which can be filled with the gas under test. The transmitted radiation intensity is rectified by a crystal, and the rectified voltage, which is modulated at 1000 cycles, is balanced against a 1000-cycle oscillator by means of an attenuator. The attenuator reading measures the absorption.

so changes the loss due to the finite resistance in the guide.[2] Attenuation beyond this figure is considered to be due to true gas absorption. A second effect is due to the change of wavelength of the radiation in the guide which changes the phase at the detector. If there is any appreciable standing wave ratio a large change of phase would make a serious error. Actually the change of phase

[2] This subject is discussed in Ramo and Whinnery's *Fields and Waves in Modern Radio*, John Wiley and Sons, 1944, p. 346.

is not great, and if the standing wave ratio is kept low the error is kept within bounds.

Beringer's results are shown in Fig. 13·2. The curve shows the variation of the attenuation constant with frequency for pure

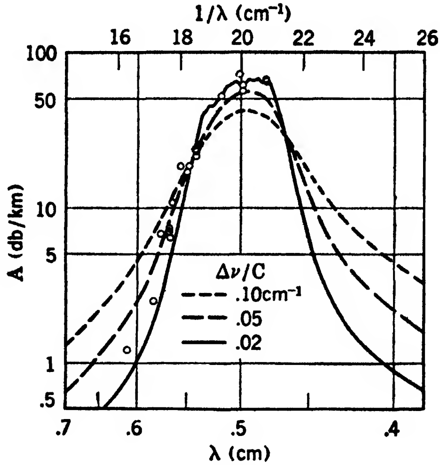


FIG. 13·2 Beringer's results for the absorption of microwave radiation by oxygen. The curve shows the variation of the attenuation constant with frequency. The solid curves are theoretical lines derived by Van Vleck for various values of $\Delta\nu/C$ which describes the natural line width.

oxygen. The whole of the absorption line is not covered, but the agreement with theory is not bad, as the fit to the curve drawn through the points shows.

Ammonia Absorption

Of great interest is the absorption by ammonia. This property was discovered by Cleeton and Williams [3] in 1934, and ammonia has been much the most thoroughly studied gas.

The pioneer work of Cleeton and Williams differs in many respects from present-day techniques. It is an astonishing achievement when one recalls the effort which went into similar endeavors eight years later during the war. Cleeton and Williams attacked the problem by standard optical methods. They relied on an echelette grating to provide radiation of a known wavelength.

[3] C. E. Cleeton and N. H. Williams, *Phys. Rev.*, **45**, 234 (1934).

The source of the radiation was a series of four split-anode magnetrons which were scaled down from 9-centimeter magnetrons built to a design by Kilgore. The oscillating system was a Lecher wire pair 4 millimeters long. The anode radius was 0.27 millimeter, and the magnetic field 11,000 gauss. Graphite anodes were used, and the oscillating system was enclosed in the vacuum envelope. With the aid of two paraboloid reflectors the radiation could be

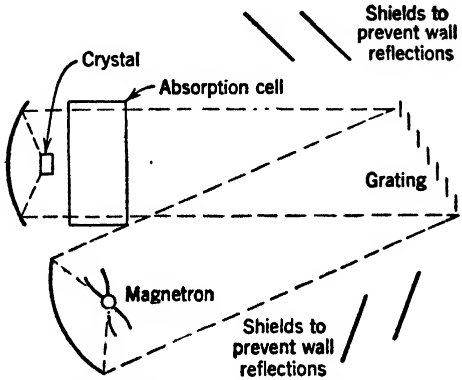


FIG. 13·3 Cleeton and Williams' apparatus for the observation of microwave absorption by ammonia. Radiation from the split-anode magnetron is selectively reflected by the grating, and the absorption in a large rubberized cloth cell is measured directly.

detected by an iron pyrite-phosphor bronze crystal, and it gave a full scale deflection on a galvanometer.

Cleeton and Williams' spectrometer is illustrated in Fig. 13·3. The magnetron source is placed at the focus of a paraboloid. The beam of radiation resulting is diffracted from an eighteen-element grating with a constant of 7.49 centimeters designed so that the elements could be adjusted to give maximum intensity at any angle in use. The diffracted beam passes through a cell of rubberized cloth 16 inches long and of rectangular section 36 by 45 inches. The radiation passing through this cell was reflected from a paraboloid onto a crystal. The rectified current operated a galvanometer.

The results obtained by Cleeton and Williams are shown in Fig. 13·4. The absorption coefficient was measured as a function of wavelength by comparing galvanometer readings with and without the cell in place and correcting for the absorption of the cell itself by taking a separate set of readings with air in the cell.

The resonance nature of the absorption is clearly seen. The line drawn is a theoretical line derived by supposing the line width to be due to the effect of collisions in modifying the spectral distribution of radiation incident on the molecules. This theory will be discussed further a little later.

With the vastly improved microwave technique resulting from war research a fresh attack could be made on the problem of the

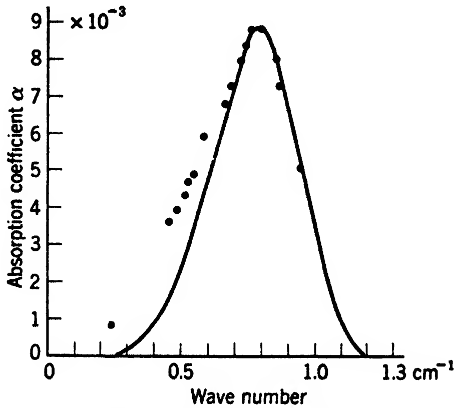


FIG. 13·4 Cleeton and Williams' original observation of the absorption of microwaves by ammonia. The curve is drawn according to the relation $\alpha = kf^2[\sin(\tau\Delta f)/\Delta f]^2$ obtained by assuming that the line width is due to the shortening of wave trains by collisions.

ammonia spectrum. This has been done by Bleaney and Penrose [4] Good,[5] and Townes.[6] The method used is essentially the same as that employed by Beringer. The microwave generator is rather easier to set up since the wavelength is around 1.25 centimeters where well-engineered tubes are available and no frequency doubling by a crystal is required. The method of indication involves sweeping the frequency of the oscillator providing the microwaves and synchronizing a sweep on an oscilloscope with it. The difference in the output of two crystals, one at the entrance to the absorbing path and the other at the end, is then displayed as a vertical deflection on the oscilloscope. The actual shape of each single absorption line is then seen on the oscilloscope screen.

[4] B. Bleaney and R. P. Penrose, *Nature*, **157**, 339 (1946).
[5] W. E. Good, *Phys. Rev.*, **69**, 539 (1946).
[6] C. H. Townes, *Phys. Rev.*, **70**, 665 (1946).

The most interesting result of this recent work is that when the pressure of the ammonia is reduced the absorption spectrum passes from the single broad band discovered by Cleeton and Williams to a series of twenty-six separate lines. The positions of these lines are shown in Fig. 13·5. They are interpreted as being due to a form of repetition of the molecular transition responsible for the absorption line itself. It is worth while to give here a brief description of the way in which the ammonia absorption arises because

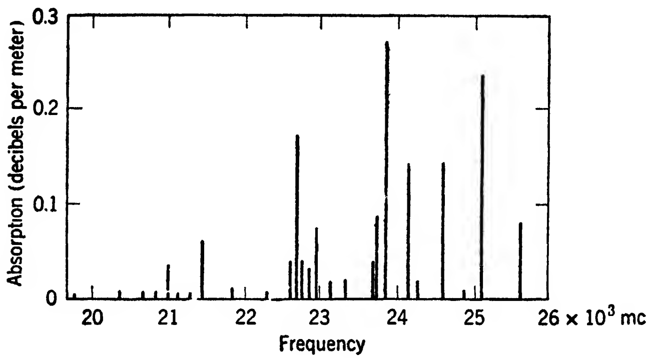


FIG. 13·5 Plot of the frequencies and relative intensities of the ammonia absorption spectrum according to Good. These lines can be analyzed in terms of the rotation levels of a symmetrical top molecule.

it is an interesting process. The ammonia molecule has three hydrogen atoms in a plane, with the nitrogen atom situated above this plane. The molecule can vibrate and also rotate. The process of vibration, in which the nitrogen atom goes back and forth through a short oscillation above the plane of the hydrogens, is responsible for a vibration spectrum having wavelengths much shorter than the microwave region. This vibration can also take place with the nitrogen atom below the plane of the hydrogens, which is an alternative configuration indistinguishable from the other. The existence of these two configurations has an effect on the nature of the energy levels of the molecule, causing them to be double. The separation of the two values corresponds to a frequency difference which is in the region of 30,000 megacycles. It is the transition between these two levels which is produced by the electric field of the microwave radiation and causes the absorption.

The doubling of the levels is analogous to the production of two frequencies when two pendulums of equal length are suspended from a common support which can itself vibrate. The two frequencies of oscillation of the pendulum are nearly equal for small "coupling," and the difference in frequency depends on the amount of coupling between the two. The analogue of the coupling of the

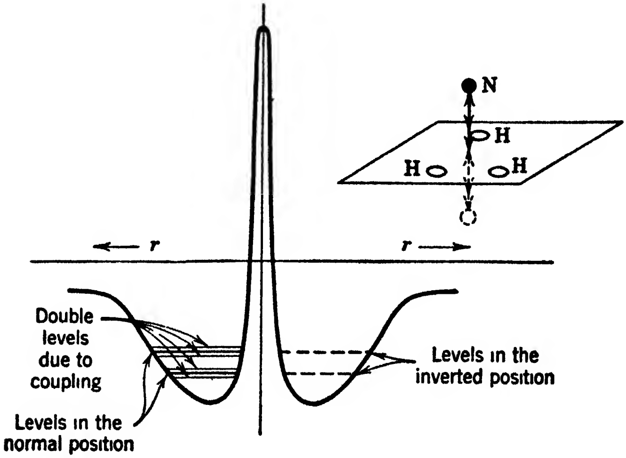


Fig. 13·6 Showing the potential energy of a nitrogen atom of ammonia in the field of the 3 hydrogen atoms. Two positions are possible, as the two valleys in the potential energy curve show. The fact that a finite chance of penetrating the plane of the hydrogens exists causes each energy level to be double. Transitions from one to the other of these double levels accounts for microwave absorption in ammonia.

pendulums is the height and thickness of the potential wall between the two configurations of the ammonia molecule. This is shown in Fig. 13·6. As the state of excitation of the molecule increases, the potential wall becomes thinner and the separation of the two levels increases. The only level concerned at 30,000 megacycles is the ground state. The reason why this does not appear as a single transition is that the whole molecule is also capable of rotation. The energy of rotation is quantized so that it has definite values. Each different value of the rotational energy modifies the diagram of Fig. 13·6 by an effect analogous to centrifugal force, so that the coupling between the two positions is different for different rotational states. This, in turn, means that the separation

between the two frequencies is slightly modified for each possible rotational state. This small change is definitely measurable, and if the frequencies of the separate absorption lines are observed the form of the potential curves of Fig. 13·6 can be plotted. Moreover, the relative intensities of the terms are informative. These vary according to a fine structure within the rotational levels, a fine structure determined by the nuclear spin of the hydrogen atoms. This is known to be $\frac{1}{2}(h/2\pi)$ and the predicted fine structure requires one absorption level in three to be double in intensity, which is found. The rotation is actually described in terms of two quantum numbers, J and K. The calculated and observed intensities for different values of J and K are in excellent agreement.

The observation of the effects of electric and magnetic fields on the inversion spectrum of ammonia has just been reported by Coles and Good. The first, known spectroscopically as the Stark effect, is caused by the polarization of the molecule by the electric field. This causes the formation of a dipole which will assume a series of definite angles of orientation. Since the dipole is associated with a considerable amount of rotating electric charge due to the normal motion of electrons, a gyroscopic action is superposed on the orientation with respect to the field. This causes precession. The rate of precession then determines the actual frequency of the absorption spectrum. Since the applied electric field has to create the dipole as well as cause the orientation, it is found that the frequency shift is proportional to the square of the field.

The effect of a magnetic field is to interact with the magnetic moment of the molecule, which is normally already present on account of the rotation of the electrons. The same result occurs, namely an orientation in the magnetic field with a precession around the magnetic field. The frequency shift is proportional to the field itself.

Carbonyl Sulfide, OCS

The OCS molecule is linear and possesses a permanent electric moment. It should therefore be possible to locate one of the transitions in its rotation spectrum, observe the microwave absorption, and study the nature of the molecule as revealed by the actual frequency of the absorption line, and the effect of an electric

field on it. In addition it should be possible to observe variations in the intensity of the components of the line when under the influence of the electric field and so make a measurement of nuclear spin. The added information from pressure broadening of the absorption renders work on this compound of some interest.

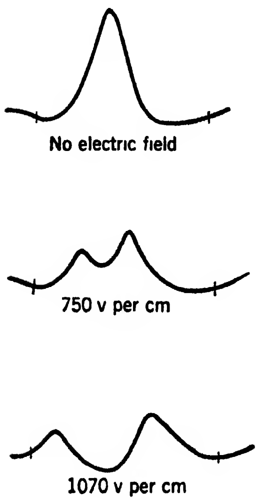


FIG. 13·7 Stark effect for a rotational line for OCS as reported by Dakin, Good and Coles. The figures represent an oscilloscope trace. The marks are 6mc apart. The line is split into two components which are shifted by different amounts.

Dakin, Good, and Coles[7] have recently reported microwave absorption at 24,320 megacycles at pressures less than one tenth of a millimeter of mercury. The measured moment of inertia is 1.379×10^{-38} gram cm^2 which agrees excellently with the calculated value of 1.38. The same authors observed the effect of an electric field on the absorption line. This is shown in Fig. 13·7 which is taken from their paper. Fields up to 1070 volts per centimeter were applied parallel to the electric vector. The splitting of the line is readily visible. Theory requires that for this particular rotational transition from $J = 1$ to $J = 2$ the difference in frequency between the two lines is

$$\Delta f = \left(\frac{3}{20} - \frac{1}{84}\right)\frac{8\pi^2 I\mu^2}{h^3} E^2 \quad (13\cdot 2)$$

E is the electric field. Since I, the moment of inertia, is fixed by the frequency of the absorption line, as already mentioned, it is possible to measure μ, the dipole moment. The value obtained is 0.72×10^{-18} electrostatic unit, and is in agreement with other data.

Hydrogen Fine Structure

The most interesting work from the theoretical point of view is that of Lamb and Retherford[8] who have measured by an indirect

[7] T. W. Dakin, W. E. Good, and D. K. Coles, *Phys. Rev.*, **70**, 560 (1946).
[8] W. D. Lamb, Jr., and I. G. Retherford, *Phys. Rev.*, **72**, 241 (1947).

means the microwave absorption frequency corresponding to a transition in the fine structure of the hydrogen atom. The results do not fit theory, and they indicate that refinements in modern radiation theory are needed.

Pressure Effects

The most striking effect of microwave absorption is the great broadening of absorption lines by increasing pressure. The original absorption line observed by Cleeton and Williams for

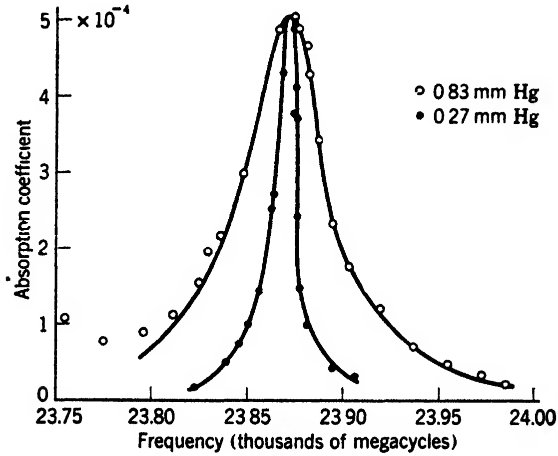


Fig. 13·8 Broadening of an absorption line of ammonia by increase of pressure. The higher pressure means more frequent collisions and hence shorter wave trains. This causes a frequency spread which appears as a broader line.

NH_3 is actually the sum of 26 lines which are completely unresolved at atmospheric pressure. The effect of increasing the pressure from 0.27 to 0.83 millimeters of mercury for one absorption line is shown in two curves due to Townes, which are reproduced as Fig. 13·8. This pressure broadening of a spectral line is due to various effects. In microwave absorption the large effect is due to the interruption of the absorption process by perturbations from other molecules, a process usually called collisions.

How this can happen is easy to see. The process of absorption has already been described. It consists of the superposition of the

microwave electromagnetic field on the electric and magnetic fields of the molecule. As a result of this superposition the molecule is in an abnormal condition, and if there is enough energy available in the light quantum of the microwave radiation there will be a change in the molecular configuration corresponding to absorption. Such a change will be very unlikely unless the incident frequency fits precisely the molecular level scheme. However, if other molecules approach the particular molecule we are considering so as to introduce an additional perturbing field it may be that the sum of the field due to the radiation and that due to the new molecule can cause a transition. This may occur even if the incident frequency is not exactly correct. At high pressures many such perturbations occur, and so the absorption process is not nearly so precisely confined to one frequency.

An estimate of the pressure effect can be obtained as follows. The processes of absorption and emission are similar. We can therefore inquire into the spread of frequency produced in an otherwise exact frequency oscillation by interruption due to collisions. If the average time between interruptions is τ seconds and the frequency for continuous emission is $f = \omega_0/2\pi$, we can apply the Fourier integral (Appendix 1) as follows. We suppose the molecule to be emitting radiation such that the electric field is zero until time $t = -\tau/2$, is then $Ae^{j\omega_0 t}$ until $t = +\tau/2$, and is thereafter zero again. If we represent the actual field by $f(t)$

$$f(t) = \int_{-\infty}^{\infty} c(\omega)e^{j\omega t}$$

where

$$c(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} f(y)e^{-j\omega y}\,dy \tag{13·3}$$

The quantity $c(\omega)$ is therefore the key to the frequency distribution. It can be evaluated by using the known fact that $f(t)$ is $Ae^{j\omega_0 t}$ while t goes from $-\tau/2$ to $\tau/2$ and is zero everywhere else. This gives

$$c(\omega) = \frac{1}{2\pi}\int_{-\tau/2}^{\tau/2} e^{jy(\omega-\omega_0)}\,dy$$

which is

$$c(\omega) = \frac{1}{\pi}\frac{\sin\frac{\tau}{2}(\omega-\omega_0)}{\omega-\omega_0} \tag{13·4}$$

This is a rather well-known result, being precisely the same as the Fourier integral for a pulse of length τ and frequency $\omega_0/2\pi$. The amplitude of a component of frequency ω is therefore given by $\sin [(\tau/2)(\omega - \omega_0)]/\omega - \omega_0$ which is the same as $\sin [(\tau/2)\ \Delta f]/\Delta f$ where Δf is the difference of frequency from the midpoint value. If we consider only the first maximum of this expression we see that it has an overall frequency width of $\pm\pi/\tau$. Most of the amplitude is confined to $\pm 2/\tau$. Thus to a simple approximation we can suppose that the line width is increased by an amount $2/\tau$ for an average interval between collisions of τ seconds.

Now if the molecules in a gas have an average speed of v centimeters per second and the average distance between collisions is L we have $\tau = L/v$. Thus we have a simple picture of the process of "pressure broadening." Using some plausible values we have $v = 10^4$ centimeters per second and $L = 10^{-3}$ centimeter, so that $\tau = 10^{-7}$ second and the broadening of the line is $2/10^{-7}$ or 20 megacycles.

This simple theory can be made more quantitative. This was done by Lorentz, and has recently been improved still more by Van Vleck and Weisskopf.[9] An accurate experimental test of the theory should in principle give information about the quantity L/v in the first place, and the lifetime of excitation of the molecule independently of collisions in the second place. Measurements of L/v can then be compared with measurements made in other ways, for example by measurement of viscosity. If there is agreement, little more need be said. However, it is unlikely that there will be agreement because the effect here considered requires only that two molecules approach near enough to supply a perturbing field which will enable an absorption to take place. The type of collision involved in the measurement of viscosity is one in which momentum is transferred from one molecule to the other. This is a much closer collision. Evidence already exists to the effect that "collision cross sections" for pressure broadening are larger than for viscosity processes. Such evidence, if it can be made quantitative enough and theory can be devised to interpret it, can give information about intermolecular forces which cannot be obtained by more usual methods.

The study of pressure effects requires highly sensitive methods of measurement of small absorption coefficients. For this purpose

[9] J. H. Van Vleck and V. W. Weisskopf, *Revs. Mod. Phys.*, **17**, 227 (1945); see also H. Margenau and W. W. Watson, *Revs. Mod. Phys.*, **8**, 22 (1936).

the waveguide technique as described above is not adequate. In several laboratories a method depending on the change in Q of a large cavity is being developed. This is advantageous because the large walls of the cavity permit large currents to flow without excessive loss, whereas the small size of a waveguide introduces loss. The problem of measurement of the field intensity inside a cavity is considerable, particularly if it is done accurately. Values of α as small as 10^{-8}cm^{-1} have been observed in this way. It seems likely that once it is solved this method of study will be very productive.

Microwave Absorption by Liquids

The study of microwave absorption by liquids has begun to yield interesting results. If a field $E = E_0 e^{j\omega t}$ is imposed on a dielectric, equation 1·13 a appears as

$$\text{curl } \mathbf{H} = \frac{1}{c}\left(\frac{4\pi}{\gamma} + j\omega k_1\right) \mathbf{E}_0 e^{j\omega t}$$

(In this equation k_1 represents the quantity $2\pi/\lambda$.) The right-hand side is accordingly complex, and the ratio $(4\pi/\gamma)/\omega k_1$ is called the tangent of the loss angle, δ. Theory indicates that

$$\tan \delta = \frac{(K + 2)^2}{K} \frac{4\pi\mu^2}{27kT} Nc \frac{\omega\tau}{1 + (\omega\tau)^2}$$

Here K is the dielectric constant, μ is the dipole moment per molecule, N the number of molecules per cubic centimeter, k Boltzmann's constant, T the absolute temperature, and τ the relaxation time of the dipole moment, meaning the time to return to normal position after a disturbing field has been removed. τ is determined by the molecular radius a, the local viscosity η, and the temperature, and simple theory [10] indicates that

$$\tau = \frac{4\pi\eta a^3}{kT}$$

Using information from other sources and assuming that η is the same as for a large scale liquid, it can be predicted that for simple molecules τ is of the order of 10^{-10} second. Now the quantity $\omega\tau/[1 + (\omega\tau)^2]$ appearing in the expression for $\tan \delta$ has a maxi-

[10] P. Debye, *Polar Molecules*, Reinhold, 1929, Chap. 5.

mum when $\omega\tau = 1$. This will accordingly be when $\omega = 10^{10}$ radians per second, or a frequency of about 1600 megacycles, which is in the microwave region. The study of the quantity $4\pi/\gamma\omega K$ or $\tan\delta$ can therefore enable the relaxation time to be measured and in addition can give a determination of the dipole moment per molecule.

The quantity $4\pi/\gamma\omega K$ is directly measured in terms of the Q of a cavity, with and without dielectric. Recalling the definition of Q as

$$2\pi \frac{\text{energy stored}}{\text{energy lost per cycle}}$$

it is seen that

$$\frac{1}{Q} = \frac{1}{2\pi} \frac{\text{energy lost per cycle}}{\text{energy stored}}$$

The energy per unit volume in a dielectric is $KE^2/8\pi$, and there is an equal amount of magnetic energy, while the power lost per unit volume per second is E^2/γ. The energy lost per cycle is then the power lost times the time of one cycle, which is $2\pi/\omega$. Thus for $1/Q$ is obtained

$$\frac{1}{Q} = \frac{\dfrac{1}{2\pi} \times \dfrac{E^2}{\gamma} \times \dfrac{2\pi}{\omega} \times \text{volume}}{2 \times \dfrac{KE^2}{8\pi} \times \text{volume}} = \frac{4\pi}{\gamma\omega K}$$

The measurement of $1/Q$ thus gives $\tan\delta$ directly. Actually the difference between the reciprocal of Q for the dielectric filling and that for vacuum filling has to be taken.

The figures for benzophenone in benzene solution are given in Table 3·1 as a sample of results obtainable in this way. They are due to Jackson and Powles.[11]

TABLE 13·1

Frequency (cycles)	Observed $\tan\delta \times 10^3$	Calculated
6.18×10^8	1.82	1.80
1.19×10^9	3.95	3.60
3.30×10^9	7.92	8.40
9.80×10^9	14.50	14.10
2.44×10^{10}	9.65	9.80

[11] W. Jackson and J. G. Powles, *Trans. Faraday Soc.*, **42A**, 104 (1946).

The calculated values are for $\tau = 16.4 \times 10^{-12}$ seconds and $\mu = 3.04$ Debye units (10^{-18} esu $\times$ cms). The agreement with theory is very good.

A study of the temperature dependence of $\tan \delta$ for one frequency has been made by Whiffen and Thompson.[12] Their results are in good agreement with theory except that the viscosity coefficient determined from the values obtained for τ is ten times too low. It therefore appears likely that the local viscosity near a polar molecule is not the same as the bulk viscosity of the whole liquid. The careful study of microwave absorption is therefore very promising with regard to research on the structure of polar liquids.

13·2 THERMAL RADIATION MEASUREMENT

It is well known that the thermal radiation from a "black body" comprises all frequencies. The distribution of the amplitude of these frequencies is given by Planck's expression (equation 13·5). This distribution depends on the temperature of the black body. Since the manner of distribution is a perfectly definite function it is only necessary to measure the amount of radiation at one frequency from an unknown temperature source and compare it with the amount from a source at known temperature to be able to measure the unknown. Such a form of radiation thermometer was described by Strong.[13] By a process of multiple selective reflection Strong obtained a narrow band of infrared radiation, for example at 8.8 microns. This was detected by a thermocouple and sensitive galvanometer.

The observation of thermal radiation in the microwave region was first made by Southworth.[14] In the course of experiments with a large antenna and a microwave receiver he observed an increase in the receiver noise when the antenna was pointed at the sun. This suggested that black body radiation from the sun was causing an added component of noise in the receiver. The equivalence between thermal noise and black body radiation had been discussed by Burgess a year or two earlier. Southworth found

[12] D. H. Whiffen and H. W. Thompson, *Trans. Faraday Soc.*, **42A**, 114 (1946).
[13] J. Strong, *J. Optical Soc. Am.*, **29**, 520 (1938).
[14] G. C. Southworth, *J. Franklin Inst.*, **239**, 285 (1945).

that the power received from the sun agreed with calculations based on the idea that the sun is a black body radiator of effective temperature 6000°K and of volume 1.4×10^{33} cubic centimeters, the energy being distributed according to Planck's expression, namely

$$U = \frac{8\pi h f^3\, df}{c^3(e^{hf/kT} - 1)} \text{ ergs per cubic centimeter} \qquad (13 \cdot 5)$$

where U is the energy density between frequencies f and $f + df$, h is Planck's constant, k is Boltzmann's constant, and c is the velocity of light. Agreement was good at 10 centimeters, but not good at about 1 centimeter.

The equivalence between thermal noise and black body radiation is interesting. Dicke [15] has pointed out a simple way of seeing that this equivalence is neccssary. If an antenna is situated inside an enclosure with walls at a definite temperature and is also connected to a line terminated by a matching resistance so that no reflections occur in the line, the power received from the enclosure must be equal to the power radiated from the noise of the resistance. If it is remembered that the antenna has a directional pattern which subtends less solid angle (as $1/\lambda^2$) as the wavelength diminishes, it is seen that the antenna selects radiation from less and less of the enclosure as the wavelength gets less. However, for very long wavelengths the Planck formula given above indicates that the radiation density increases as $1/\lambda^2$, so that the radiation absorbed by the antenna is independent of wavelength, but only if the temperature is constant. If the actual values for the antenna gain and the energy distribution are substituted numerically it is seen that the radiation collected per second by the antenna is $kT\,\Delta f$, precisely the same as the noise power of the resistance.

Microwave Radiometer

An interesting and well-designed microwave radiometer has been described by Dicke in the paper already cited. The block diagram is shown in Fig. 13·9. An antenna picks up radiation of all frequencies within its bandpass. A selection of these is made

[15] R. H. Dicke, *Rev. Sci. Instruments*, 17, 268 (1946).

by means of a local oscillator and a wideband intermediate frequency amplifier. The black body radiation appears as noise which produces an average positive voltage after rectification of the intermediate frequency. This voltage measures the intensity of the radiation in the appropriate frequency band. Such a simple

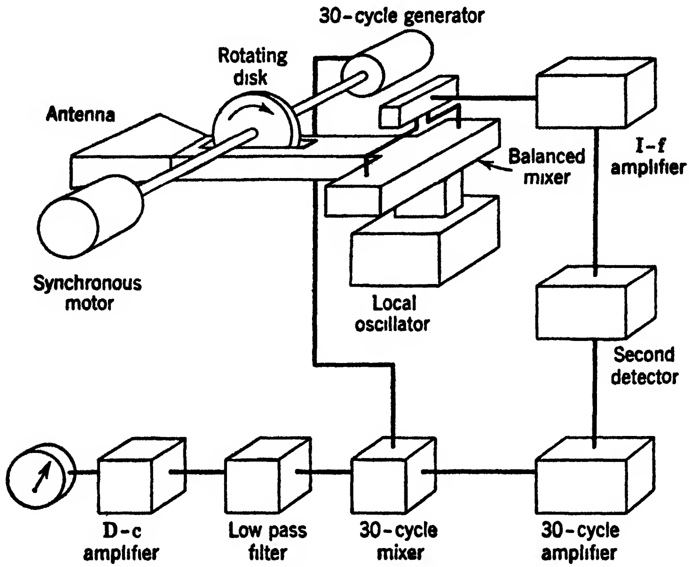


FIG. 13·9 Schematic diagram of Dicke's microwave radiometer. The antenna picks up black body radiation which is amplified and observed as increased receiver noise. To eliminate drift effects the noise from the antenna is compared with the noise from an absorbing strip 30 times per second. This is done by rotating a disk in a slot in the guide and amplifying the 30-cycle modulation introduced if the antenna noise differs from the absorbing strip noise. A balanced mixer minimizes effects from mismatch due to the rotating disk.

device will work, but it is not very sensitive. In order to increase the sensitivity the bandwidth must be wide, small fluctuations in line voltage and the like must be eliminated, and either a perfect impedance match to the antenna must be secured or a mixer must be so designed that the effects of a mismatch are not serious.

These three design considerations are met by Dicke as follows. First, the i-f bandwidth is made as large as possible by good technique. In his radiometer the figure was 8 megacycles. This

could be improved today by a factor of three or so. Second, in order to reduce the effect of line voltage variation or temperature drifts the instrument is constructed to compare very rapidly the noise from the antenna with the noise from an absorber placed in the waveguide. This is done by inserting a rotating wheel of absorbing material in a slot in the guide. The wheel is cut asymmetrically so that as it rotates it is in the guide for half the rotation and out of the guide for the other half. The rotation effects the change 30 times a second. The existence of a difference between the noise from the antenna and that from the wheel will introduce a 30-cycle component into the output of the receiver. This component is all that is of interest; any effects due to line voltages will presumably not have this precise frequency, and therefore if the amplifier is designed to record only the 30-cycle component a stable receiver of great sensitivity is available. This 30-cycle sensitivity is achieved by rectifying the output of the wideband intermediate frequency amplifier and sending it through a narrow band 30-cycle amplifier. In addition the amplified 30-cycle voltage is reduced to zero frequency by beating it with the output of an a-c generator mounted on the same shaft as the rotating wheel. This then takes care of any slight departures from precisely 30 cycles. The output is observed on a sensitive voltage-measuring device. In principle this can be a sensitive galvanometer but, like most physicists, Dicke is opposed to temperamental suspensions, and he substitutes a low pass filter, a d-c amplifier, and a microammeter.

The introduction of this very powerful method of comparison of noise effects has some repercussions. If the rotation of the wheel causes a variation in the matching of the r-f line to the mixer there will be 30-cycle fluctuations in the i-f output from the mixer. To eliminate these one deliberately selects the beat voltage from crystals placed at high and low points in the standing wave pattern. By taking the difference between these the effect of a mismatch is greatly reduced. This kind of mixer can very simply be designed into a magic tee (see Chapter 4). The local oscillator is introduced into the E plane branch, while the signal is fed into the H plane branch. The two crystals are placed in the two arms, one of which is an eighth wavelength longer than the other. The i-f output is then transformer-coupled to the i-f amplifier. The operation of a balanced mixer of this type is described more fully in Chapter 8.

The whole instrument is calibrated by placing an absorbing load at known temperature in place of the antenna. The meter reading can then be converted to absolute temperature directly. The instrument will detect 0.5°C and with a 16-megacycle bandwidth the corresponding power is 10^{-16} watt.

Not a great deal of work has been done with this very interesting device. Suggestions for its use appear in the literature, ranging from measurements of very low temperatures to the observation of "cosmic noise." During the war it was considered important to know the nature of atmospheric absorption for microwaves in the region of 1 centimeter, and accordingly members of the Advanced Development Group at the Radiation Laboratory made such measurements. These are reported by Dicke, Beringer, Kyhl and Vane.[16] This use of the radiometer is somewhat indirect. A fairly simple analysis shows that, if an antenna is embedded in a medium at a temperature T and of length l bounded by another medium at zero temperature, the antenna temperature t is

$$t = T(1 - e^{-\alpha l}) \tag{13·6}$$

where α is the absorption coefficient. The measurement of t for a known atmospheric temperature T therefore enables the value of the atmospheric absorption coefficient α to be measured. Actually the earth's atmosphere is not so simple a medium. Nevertheless by varying the angle of the antenna from the zenith and making antenna temperature measurements it is possible to fit the data reasonably well with a fixed absorption coefficient (at a single wavelength) for the water vapor distribution in the atmosphere. In this way the values 0.011, 0.026, and 0.014 decibels per kilometer per gram of water per cubic meter were found for 1.00, 1.25, and 1.50 centimeter wavelengths respectively.

In the course of this work it was found that the noise from space beyond the atmosphere is very small. This is interesting because considerable noise is observed at much longer wavelengths and is apparently associated with the Milky Way.

A few exceedingly interesting measurements have been made by Dicke and Beringer[17] on radiation from the sun and the moon. They confirmed Southworth's observation of microwave radiation

[16] R. H. Dicke, R. Beringer, R. L. Khyl, and A. B. Vane, *Phys. Rev.*, **70**, 340 (1946).

[17] R. H. Dicke and R. Beringer, *Astrophys. J.*, **103**, 375 (1946).

from the sun and found an effective temperature at these frequencies of 10,000°K, which is higher than the 6000° found from the usual solar radiation measurements. By observation of the radiation from the sun during a partial eclipse they showed that the size of the sun's disk at 1.25 centimeters is the same as seen optically. In a later experiment they observed that the temperature of the surface of the moon is 292°K. Dr. Beringer has pointed out to the authors that the observation of the rate of cooling of the moon's surface during an eclipse would be most interesting as it would give some information about the material content of the moon.

We shall end this section with a quotation from Southworth's paper. "Perhaps there lies ahead a new field of astronomical and terrestrial research comparable in interest and scope with that that has prevailed during the last two decades with respect to the ionosphere. The sun being a localized source of radio waves situated entirely outside the earth's atmosphere provides, for the first time, angles of attack on the structure of the earth's atmosphere not previously possible."

13·3 MICROWAVE PARTICLE ACCELERATORS

Multiple acceleration methods for speeding up electrons, protons, deuterons, or other nuclei have been proposed, and devices have been constructed several times during the past two decades. The cyclotron is the most successful of those in operation to date. There has always been the question, however, regarding the complication introduced by a magnetic field when the particle approaches very high energies so that its mass is increasing considerably. The recent introduction of frequency modulation into cyclotron technique has answered the question very effectively as regards energies as high as 200 million electron volts (Mev) and probably much higher. Even so there are advantages to an accelerator in which the particles move in straight lines, notably the ready accessibility of the beam, and there has been much speculation about the possibility of using the high power available at microwave frequencies to develop high voltages for such an instrument.

Before such a linear accelerator is considered, it is worth while to review some possibilities for developing voltages for accelera-

tion. One such was suggested by Collins, who proposed to use a waveguide bent many times back and forth with a hole through all the folds of the guide. Power would be fed in from a magnetron, which would develop a high voltage across the guide and accelerate a charged particle exposed to the voltage. By correct spacing of the folds the particle could be accelerated each time it reached a new fold of the guide. This method should work but it seems to require large power.

A second proposal was made by Schwinger. A single cavity is excited by high power. It is placed in a magnetic field, and electrons are introduced into the cavity. These will be accelerated by the high electric field in the cavity and will pass out of an opening on the other side at high velocity. They then execute a circular motion in the magnetic field and return to the first opening in the cavity in time to be accelerated again. This process goes on, the electrons following ever widening circles until their energy is very great. The difficulty of this method is the introduction of the electrons in the first place, a difficulty so great that so far no one has started to construct such a machine.

A third proposal, made in a seminar at Harvard by Getting, and independently by Hudspeth,[18] is to use a spiral of waveguide and to send the particles to be accelerated across a diameter of the spiral. Advantage is then taken of the geometrical fact that each turn of the spiral increases the length of the guide by a constant amount. By adjusting the wavelength to match this amount it is possible to arrange for acceleration every time the particle crosses the guide. This seems a reasonable proposal and probably would work, but would get very formidable at high energies.

Cavity Accelerators

The method which at present seems most practical is the use of a series of cavities excited to as high power as is available and spaced so that the particle uses up the repulsive time outside the cavities and is in the accelerating field while it traverses the cavities.

Three major problems have to be solved to make such an instrument work. The first is the development of a high voltage across the cavity. The second is some sort of control over the phase of

[18] E. L. Hudspeth, *Phys. Rev.*, 69, 671 (1946).

the electric field in each of the cavities. The third is the focusing of the beam. Of these it can be said that the first appears to be satisfactory. The second would seem to be possible, and the third is still in the speculative stage, with opinions being developed on the subject. The operation of a several-stage accelerator giving more than 10 Mev has been achieved but it is still early to appraise the results.

We can consider these three problems briefly and then describe a typical proposal for a multistage accelerator.

The voltage developed across a cylindrical cavity can be calculated roughly without much trouble. A modern magnetron can deliver 1 megawatt at 10 centimeters. A cylindrical cavity which is well constructed has a shunt impedance of 10^6 ohms or so. The relation *power* $= E^2/R$ gives the value 10^6 for the voltage across the cavity. Therefore the gain in energy per stage is high. It now remains to make an engineering decision about the shape and dimensions of the cavity used. The highest voltages can be developed across cylindrical cavities which are relatively long. On the other hand it may pay to get a little less voltage per cavity and use more of them, thus obtaining a greater acceleration per foot. This is largely up to the individual designer. The lowest mode of a cylindrical cavity is excited when the radius a and wavelength λ are related by $\lambda = 2.61a$. This means for 10-centimeter radiation a radius of about 4 centimeters. An inspection of equation 2·25 shows that the shunt resistance R_s is proportional to Area $\times f/\alpha$. (α is the skin depth, f the frequency.) Then, other things being equal, the area increases linearly with the length of the cavity and so does the shunt resistance. Hence the voltage for a given power rises as the square root of the length. In order to maintain a simple oscillation in the cavity it is probably wise to keep the length to the order of the wavelength, in which case the shunt resistance as given by Table 1·1 is of the order of 20 megohms, and the voltage developed for a megawatt input is 4.5 million. This is probably higher than can be achieved in practice. If such cavities are spaced a wavelength apart the actual acceleration achieved is about 7 Mev per foot. An accelerator 100 feet long could then, in principle, produce particles of energy 700 Mev, which is far beyond any energy attained up to the present. The power required for such a machine would be 150 megawatts. This is peak power. The average power

would be about 150 kilowatts, and if the efficiency were 20 per cent this would be a total power need of 800 kilowatts. This is high but not by any means impossible.

The question of the phase control of the cavity voltage is more serious. There are two approaches to this. The first is to make all the cavities parts of a single tank so that the walls of each cavity are shared with adjoining cavities and with the whole tank. Now if power is fed in from a series of separate oscillator tubes with a provision for matching each tube to the tank and a tuning

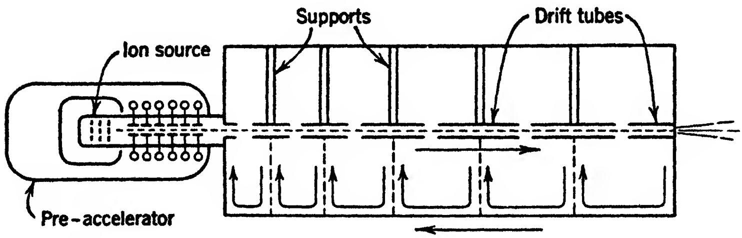


FIG. 13·10 Schematic diagram of Alvarez' linear accelerator. Protons are injected at 4 million volts from an electrostatic accelerator, and then pass through a succession of field free drift tubes and accelerating gaps where the radiofrequency field is in the direction of travel of the particles.

adjustment on each, the fact that the whole tank is effectively being driven in a single high mode of oscillation will require that definite phase relationships be established along its length. Such a tank is being constructed by Alvarez, Panofski, and their group at the University of California.[19] It is illustrated schematically in Fig. 13·10. The dotted lines represent the hypothetical walls of a series of re-entrant cavities. The re-entrant parts form tubes separating gaps, the gaps are the points where the field does the accelerating, and the tubes, called "drift tubes," are for the purpose of protecting the particle from the decelerating phase of the oscillation. It will be seen that the currents in the wall adjoining two of the hypothetical cavities are opposed. Hence it is reasonable to dispense with them and construct the tank as a unit with the drift tubes supported down the axis of the cylinder. The power from a large number of separate oscillators is fed in. Provision is made for some matching adjustment in each line.

[19] See three abstracts in *Phys. Rev.*, 70, 447 (1946).

Note that each drift tube increases in length as the proton becomes more energetic. In order to avoid the necessity of very short tubes initially, the protons are injected from an "ion source" consisting of a pressure electrostatic accelerator which gets them up to 4 Mev before the linear accelerator takes over.

A different approach to the phase control problem is the use of a master oscillator which drives all the other oscillators in the same phase. It is then possible to adjust the phase of any one cavity at will by varying the length of the line which feeds the cavity. This is a very direct and simple approach but is somewhat difficult to achieve at microwave frequencies since magnetrons are not readily driven. At a wavelength of 50 centimeters high power triodes are available. Schultz and his group at Yale are employing this wavelength for an accelerator which uses this method of phase control.

The kind of particle to be accelerated dictates to some extent the desirable frequency. For protons, which move relatively slowly and do not approach the velocity of light until their energy is in the billion-volt region, there is much to be said for longer wavelengths with larger cavities. For electrons which rapidly reach a velocity very close to that of light and thereafter simply gain in mass, the drift tubes or their equivalent can be shorter, and short wavelengths are attractive. At present work is going ahead at 200, 600, and 3000 megacycles.

The third problem, that of focusing the ions while they are accelerated, is the most serious. The nature of the acceleration by an alternating field is different from that of a d-c field. In the latter case it is possible to use the action of an electrostatic lens at the gap between two electrodes to give strong focusing. This depends fundamentally on the fact that as the particle traverses the gap the curved lines of force initially pull it to the center and then, as it traverses the second half of the gap, pull it to the edge. The second half is relatively ineffective because the particle is going faster and is harder to deflect. In the alternating case the field distribution is quite different, and moreover is varying with time so that there is no such lens action. In order to overcome this difficulty Alvarez has proposed closing the openings of the drift tubes with thin foils of beryllium. This has the effect of restoring the lens action since the field is now that between two planes and not that between two open cylinders.

Small currents of 30 million electron volt protons have been reported with this accelerator but a full experimental appraisal is not available at the time of writing. Whether this device can compete with the frequency-modulated cyclotron in the field of high energy nuclear physics is doubtful.

13·4 MICROWAVES AND SUPERCONDUCTION

The fact that lead, tin, and a number of other elements become superconducting at temperatures close to absolute zero is well known. Recent work has also shown that superconduction is suppressed by a magnetic field in the sense that a lower temperature is needed to cause superconduction if a magnetic field is applied. Also it has been shown that superconduction is heavily concerned with the surface layers of the superconductor. There is no established theory of superconduction, and it is one of the most intriguing branches of modern physics.

The facts that superconduction is concerned with thin surface layers and that a magnetic field influences it at once suggest a study of the effect of microwave radiation on a superconductor. The microwave currents are confined to a very thin layer on account of the "skin effect" and carry with them large alternating magnetic fields. Some experiments along these lines have been reported by London,[20] who showed that at 20.5 centimeters the radiofrequency resistance changed *gradually*, as the temperature of a sample of tin was cooled below the transition temperature. The normal resistance changes *abruptly* by an enormous factor so the r-f resistance is certainly different.

There is no explanation of this effect. It has been suggested by London that the results fit a hypothesis to the effect that there are present both "normal" and "superconducting" electrons in a metal below the transition point. For d-c measurements the superconducting electrons play the important part. However, at radiofrequencies the magnetic field due to the changing electron current penetrates into the conductor and there causes motion of the normal electrons with an associated normal resistance. According to London all the electrons available for conduction are superconducting at absolute zero but the number diminishes with tem-

[20] H. London, *Proc. Roy. Soc.*, 176A, 522 (1940).

perature according to a relation

$$\frac{N_s}{N_0} = 1 - \frac{3}{4}\left(\frac{T}{T_t}\right)^2 - \frac{1}{4}\left(\frac{T}{T_t}\right)^{16} \tag{13·7}$$

where N_s is the number of superconducting electrons, N_0 the number of ordinary electrons, T the temperature in degrees Kelvin, and T_t the transition temperature. This relation is empirically derived from measurements on thin films and colloids made by Appleyard, Bristow, London, and Misener, and also by Shoenberg. It can be seen that the r-f resistance will therefore change gradually, in a manner determined by N_s/N_0, from a very small value near 0°K to a normal value at the transition temperature. London used an ellipsoid of tin surrounded by a resonator shaped like a cylinder with slots cut at the ends. Power was fed in from a split-anode magnetron and the heat evolved due to r-f currents measured by observing the rate of evolution of helium gas from helium liquid. For the same r-f power, measured at the input, the rate of evolution of gas was determined as the temperature changed. The result, expressed in terms of r-f resistance, is as already described.

London's work has recently been expanded for tin and mercury by Pippard.[21] Pippard used a different measurement technique. He used the fact that, as the skin depth changes because of changing resistance, the effective size of a resonant cavity changes slightly and hence the resonant frequency changes. By studying the change of resonant frequency as the resonant cavity (in this case a parallel wire resonator) is reduced in temperature, Pippard determined the way in which the skin depth changes with temperature. The results are shown in Fig. 13·11, which is taken from Pippard's article. The change in penetration depth starts rapidly but continuously at the transition temperature and thereafter proceeds more slowly until only a gradual change occurs.

A very interesting feature of this research has been the study of the skin effect in normal conductors at low temperatures. The resistance of a conductor diminishes as the temperature diminishes; this means that the chance of collision between an electron and an atom becomes less and the average velocity therefore greater because the electron is accelerated for a longer time. It will be recalled from the treatment on page 19 that the conductivity con-

[21] A. B. Pippard, *Nature*, **159**, 435 (1947); *Proc. Roy. Soc.*, **A191**, 370 et seq.

sists of the repeated acceleration and stoppage of electrons. If the chance of collision is less, the mean free path becomes greater until it may exceed the skin depth as predicted by usual theory. When this occurs there is clearly need for modification of the theory because the electron will certainly not stop until it makes a collision, and therefore the penetration of fields into the conductor will be different from the prediction. Pippard has found that the skin

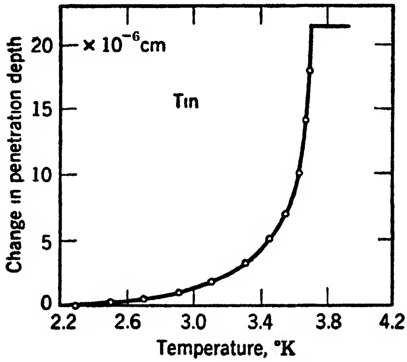


FIG. 13·11 Pippard's curve showing the way in which the penetration depth, as deduced from the change in the resonant frequency of a parallel wire resonator, varies below the superconducting transition temperature. The change is gradual, perhaps indicating a slow increase in the fraction of superconducting electrons as the temperature drops.

depth does not decrease with temperature as expected according to theory below about 20°K. This is linked with the large mean free path, and a simple treatment based on this fits the experiments reasonably well.

One apparently simple experiment is to construct a cavity out of a superconductor and to measure its Q as the temperature falls below the transition point. This has been tried by Bitter, Garrison, Halpern, Maxwell, Slater, and Squire at M.I.T. and independently at Yale by W. Fairbank. The experiment consists in feeding power from a klystron, controlled by a Pound stabilizer, as described in Chapter 12, into a lead cavity in a helium refrigerator. The standing wave ratio in the input line was measured as a function of frequency, and in two runs the Q was observed to change from a value of 7000 to a value of 10^6 somewhere just below the transition point. More complete work should be of great interest.

There is one general reason why microwaves and low temperature experiments should be revealing. This is that the thermal energy of a component of motion (degree of freedom) at 1°K is $\frac{1}{2}k$ or 0.77×10^{-16} erg. This is the same as the energy of a quantum of radiation of frequency 11,800 megacycles or of wavelength 2.54 centimeters. The absorption of such a quantum at low temperatures therefore produces a large effect and should yield experimental results of value.

The actual combination of microwaves and low temperatures involves the use of two rather elaborate techniques. So far this has held the work down to a rather small output. As the experimental mastery of the combination is attained, a considerable increase of interesting results can be expected.

13·5 OTHER FIELDS OF APPLICATION

We can mention briefly here one or two other fields of application. Of great importance is the study of dielectrics at microwave frequencies made by von Hippel. An extensive report on this is to appear in the *Radiation Laboratory Technical Series.* The method is described in an article by Roberts and von Hippel.[22] Another research field which is related in character to microwave techniques is the method of studying resonance with atomic and nuclear precession in a magnetic field. This was devised by Purcell, Torrey, and Pound [23] and has been applied to the study of nuclear magnetic moments and to the Brownian motion of rotation of molecules.

Conclusion

In this chapter we have described briefly four lines of research involving microwaves. Many more exist. We cannot hope to do justice to them. We can only point out the possibilities, outline the techniques, and hope that we have provided some aid to the ingenuity of the reader.

[22] S. Roberts and A. von Hippel, *J. Applied Phys.*, **17**, 610 (1946).
[23] E. M. Purcell, H. C. Torrey, and R. V. Pound, *Phys. Rev.*, **69**, 37 (1946).

APPENDIX 1

THE FOURIER INTEGRAL

The Fourier integral is a means of transformation which can be applied to any function $f(x)$ which fulfils the following three conditions:

1. $f(x)$ and its first derivative are continuous in any finite interval except for the existence of a finite number of discontinuities.

2. At such discontinuities the value of the function is the arithmetic mean of the upper and lower values.

3. $\int_{-\infty}^{\infty} |f(x)|\,dx$ is finite.

When these conditions apply,

$$f(x) = \int_{-\infty}^{\infty} c(\omega)e^{j\omega x}\,d\omega$$

where

$$c(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} f(y)e^{-j\omega y}\,dy$$

This combined expression is called the *Fourier integral.* The function $c(w)$ is called the *Fourier transform* of $f(x)$.

The proof of the theorem is not difficult and is given in many standard mathematical texts. The conditions, although very restrictive mathematically, can be seen to apply to any but the most abstract physical functions, and certainly to the processes considered in this book. It can be seen that $f(x)$ is represented in terms of two new quantities. The first is a periodic function of x, controlled by a frequency constant ω, which can be identified

with a frequency of oscillation of some kind. The second is an amplitude function $c(\omega)$ of the frequency constant. The representation requires integration over all values of ω. This is a mathematical statement of the fact that $f(x)$ can be represented by a sum of terms of different frequencies with properly chosen amplitudes. The choice of the amplitudes is governed by $c(\omega)$, which accordingly controls the *frequency spectrum*. The form of $c(\omega)$ is given by the second integral, and is in many cases remarkably simple. Where this is true the theorem is of great use.

Three major applications of this theorem are made in this book. It is applied to a pulse of fixed frequency but finite time (i.e., a pulse of r-f) and to a single pulse in time (i.e., a "video" pulse), and it can be used to relate the secondary pattern of an antenna with the primary pattern of the feed.

The simplest application is to the pulse of fixed frequency (angular frequency ω_0), the transmitted pulse of a radar. Here we know that the variable is time t, and explicitly $f(t)$ is an oscillatory function $Ae^{j\omega_0 t}$ during the time from $-\tau/2$ to $+\tau/2$ and zero at all other times. This represents, for example, the output of a magnetron to which a pulse of duration τ seconds is applied, and the result should bear comparison with the appearance of a spectrum analyzer. We have

$$f(t) = \int_{-\infty}^{\infty} c(\omega)e^{j\omega t}\,d\omega$$

where

$$c(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} f(y)e^{-j\omega y}\,dy$$

Using the known form of $f(y)$ we can evaluate $c(\omega)$. The result is seen to be

$$\begin{aligned}
c(\omega) &= \frac{1}{2\pi}\int_{-\tau/2}^{\tau/2} Ae^{j\omega_0 y}\,e^{-j\omega y}\,dy \\
&= \frac{A}{\pi(\omega_0-\omega)}\left[\frac{e^{j(\omega_0-\omega)(\tau/2)} - e^{-j(\omega_0-\omega)(\tau/2)}}{2j}\right] \\
&= \frac{\tau A}{2\pi}\,\frac{\sin\dfrac{\tau}{2}(\omega_0-\omega)}{\dfrac{\tau}{2}(\omega_0-\omega)}
\end{aligned}$$

This result shows that the frequency spectrum of such a pulse has a maximum where $\omega_0 = \omega$; or, on the correct frequency, a series of minima where $(\tau/2)(\omega_0 - \omega) = n\pi$, n being an integer; and a series of subsidiary maxima where $(\tau/2)(\omega_0 - \omega) = [n + (3/2)]\pi$. These maxima fall off in amplitude proportionally to $1/(\omega_0 - \omega)$. The amplitude of $c(\omega)$ drops to half its maximum value when $\dfrac{\sin \frac{\tau}{2}(\omega_0 - \omega)}{\frac{\tau}{2}(\omega_0 - \omega)} = \dfrac{1}{2}$. This occurs when $\omega_0 - \omega = 3.7/\tau$. Since $f = \omega/2\pi$, the frequency difference to half maximum is $3.7/2\pi\tau$, or very nearly $1/2\tau$.

Very similar to this is the case of a simple voltage pulse extending from time $-\tau/2$ to $+\tau/2$. Here $f(t)$ is A during this time and zero at all other times. We have again

$$f(t) = \int_{-\infty}^{\infty} c(\omega)e^{j\omega t}\, d\omega$$

where

$$c(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} f(y)e^{-j\omega y}\, dy$$

Evaluating $c(\omega)$ as before and using the fact that $f(y)$ is now A when $|y| < \tau/2$ and zero elsewhere gives

$$c(\omega) = \frac{1}{2\pi}\int_{-\tau/2}^{\tau/2} Ae^{-j\omega y}\, dy$$

$$= \frac{\tau A}{2\pi}\left[\frac{\sin \omega \frac{\tau}{2}}{\frac{\omega\tau}{2}}\right]$$

This is a frequency spectrum having a maximum when $\omega = 0$ and a series of secondary maxima and minima where $(\tau/2)\omega = [n + (3/2)]\pi$ and $n\pi$. The frequency spread to half maximum amplitude is, as before, roughly $\frac{1}{2}\tau$.

The third application is to the form of the secondary pattern from an illuminated aperture. We can take a very simple case, that of a slit extending from $-a/2$ to $+a/2$. We have first to

consider a little physics. At a point distant r from the center of the slit the amplitude is the sum of effects from the whole of the slit. On the far side the phase is increased by $[(x \sin \theta)/\lambda]2\pi$ where x is the distance from the center of the slit, θ the angle made by the direction of r with the normal to the slit, and λ is the wavelength. On the near side it is diminished by a like amount. The disturbance at r due to a section dx of slit at x is therefore given by

$$\frac{A(x)}{r^2} e^{j\left[\frac{2\pi}{\lambda}r + \left(\frac{x \sin \theta}{\lambda}\right)2\pi\right]} dx \quad \text{or} \quad \frac{A(x)}{r^2} e^{j\frac{2\pi r}{\lambda}} \cdot e^{j(x \sin \theta)\frac{2\pi}{\lambda}} dx$$

where $A(x)$ is the primary illumination function describing how the intensity of illumination is distributed over the slit. This can be divided into $(1/r^2)e^{j\frac{2\pi r}{\lambda}}$ expressing the distance of the whole slit from the point considered, and $A(x)e^{jux}\,dx$, where $u = (2\pi \sin \theta)/\lambda$ describing the effect of distribution over the slit. Treating the latter part alone we have for the amplitude at distance r along direction θ

$$f(u) = \int_{-a/2}^{a/2} A(x)e^{jux}\,dx$$

Now we can evaluate $f(u)$ if we know $A(x)$. We can also apply the Fourier transform and obtain

$$A(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(u)e^{-jux}\,du$$

which enables us to see what primary pattern $A(x)$ is needed to produce a required secondary pattern $f(u)$. This relationship is of great value in predicting the correct illumination for given sidelobe amplitudes.

The Fourier integral is clearly related to the Fourier series. The Fourier series applies to a repetitive process, and actually is quite valid for most of the purposes of frequency analysis of radar pulses. The relationship between the two is excellently shown in Guillemin's very complete work, *Communication Networks.*[1]

[1] John Wiley and Sons, Inc., Vol. I, 1931, Vol. II, 1935. The reader is referred to the chapter on the Fourier integral for much physical insight.

One important and elementary consequence of Fourier analysis is worth mentioning. The fact that any "physical" function of time can be analyzed into a frequency spectrum justifies circuit analysis in terms of single frequencies, for then the summation of these can be used to describe any applied current or voltage variation. The familiar analysis in terms of resistance R and reactance $1/j\omega C$ or $j\omega L$ is therefore able to describe the behavior of a circuit dealing with pulses.

APPENDIX 2

CURL AND STOKES' THEOREM

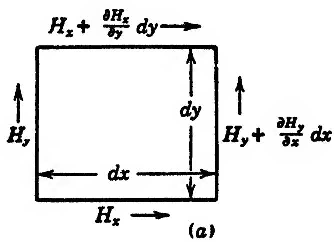
The operation of the electric and magnetic fields, a basic property of Nature, turns out to be as follows. An electric field which is changing with time causes a magnetic field, but in such a way that only the part of the magnetic field which contributes to a rotation is concerned. Therefore it is necessary to see just what part actually does contribute to a circular rotation. To see this consider the rectangle of sides dx, dy as shown in Fig. A·1. Suppose that H has components H_x, H_y, which increase to $H_x + \dfrac{\partial H_x}{\partial y}\,dy$ and $H_y + \dfrac{\partial H_y}{\partial x}\,dx$. Now proceed around the rectangle in a clockwise fashion. It is very simple to check that the net H encountered in going around is $\dfrac{\partial H_x}{\partial y}\,dy - \dfrac{\partial H_y}{\partial x}\,dx$. The quantity which determines whether this be large or small is therefore $\dfrac{\partial H_x}{\partial y} - \dfrac{\partial H_y}{\partial x}$. Now to return to the process which causes this net rotational effect: it is recalled that the electric field responsible is along the z axis (that is, perpendicular to the paper). Therefore the quantity $\dfrac{\partial H_x}{\partial y} - \dfrac{\partial H_y}{\partial x}$ must also be related to the z axis. This is done by asserting that a vector, determined by the rotational part of H exists, directed along the z axis and of magnitude $\dfrac{\partial H_x}{\partial y} - \dfrac{\partial H_y}{\partial x}$. It is a component of curl **H**. Since there is no reason to suppose rotation can occur in the xy plane only, the xz

and yz planes must also be considered. Each of these defines a *component* of the new vector, and the vector from which these are derived clearly describes all possible rotation in H. This is then defined as curl $\mathbf{H}$. If three unit vectors $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ are used along the x, y, and z axes respectively, curl $\mathbf{H}$ is defined as

$$\mathbf{i}\left(\frac{\partial H_y}{\partial z} - \frac{\partial H_z}{\partial y}\right) + \mathbf{j}\left(\frac{\partial H_z}{\partial x} - \frac{\partial H_x}{\partial z}\right) + \mathbf{k}\left(\frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y}\right).$$

The process of vector addition which applies here, it may be remarked, is the process of finding a resultant.

Stokes' theorem expresses a very simple concept, though in rather unusual language. In Fig. A·1 (*b*) a somewhat odd-shaped



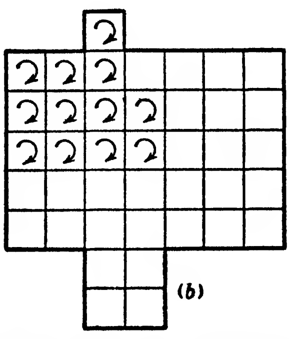


FIG. A·1 (*a*) One component of curl; (*b*) illustrating Stokes' theorem, which expresses the fact that clockwise traversal of each small rectangle leaves only the boundary traversed in one direction.

outline is divided into small rectangles. The concept expressed by Stokes' theorem is the almost self-apparent fact that, if every rectangle is traversed clockwise, *all but the outside line* is traversed twice, once in each direction. Hence such a process can be de-

scribed either as the summation of many circles or as the passage once clockwise around the whole figure. If the rectangles are made very small the process can be applied to a figure of any shape to a good degree of approximation.

Now consider the *line integral of H* around one small rectangle. By this is meant the algebraic sum of the products such as $H_x\,dx$, namely the component along one side times the length of the side. For one rectangle this is

$$H_y\,dy + H_x\,dx + \frac{\partial H_x}{\partial y}\,dy\,dx - H_y\,dy - \frac{\partial H_y}{\partial x}\,dx\,dy - H_x\,dx$$

$$= \left(\frac{\partial H_x}{\partial y} - \frac{\partial H_y}{\partial x}\right)dy\,dx = \text{curl}_z\,H\,dy\,dz$$

where $\text{curl}_z\,H$ is the z component of curl $\mathbf{H}$.

For this special case therefore we have

$$\oint H\cos\theta\,dl = \int_{\text{surface}} |\,\text{curl}\,\mathbf{H}\,|\cos\phi\,dS$$

where $H\cos\theta$ denotes the component of H along dl and $|\,\text{curl}\,\mathbf{H}\,|$ $\cos\phi$ denotes the component of curl $\mathbf{H}$ perpendicular to the area dS. θ is the angle between H and dl, and ϕ is the angle between curl $\mathbf{H}$ and the normal to the area dS.

Now, extending the process to all the small integrals, we can assert that the left-hand side adds up only as far as the outside edge is concerned, while the right-hand side simply adds up over the area, so that the integral signs can now be held to refer to the whole loop and surface respectively. This is Stokes' theorem.

This simple explanation may serve in place of a proof. The theorem holds even if the area is not chosen perpendicularly to the z direction.

For a very clear account of both curl and Stokes' theorem see Skilling's *Fundamentals of Electric Waves* (John Wiley and Sons, 1942). Skilling introduces the idea of curl meters, which are little paddle wheels carrying a right-hand screw thread on their axis. These are inserted in any spot where curl is to be investigated. The tendency to move the screw is a measure of the curl. For someone who needs to visualize a process this is the best way to grasp the idea of curl that the authors know.

APPENDIX 3

UNITS

"Anyway we shall not fall into the common error of becoming slave to a particular unit system and shall not hesitate to change the units whenever it is advantageous to do so."—E. U. CONDON.

In order to show most clearly the fundamental properties of electric and magnetic fields and their causes, the system of units which shows this most simply has been chosen. Very rarely is a comparable system of units adopted in practical work; witness the fact that temperature is not ordinarily given in degrees Kelvin. A practical system of units has been developed in the past fifteen years which combines many advantages in practice with a reasonably simple foundation. Since we are not claiming that the system of units we have used has any special virtue for practical use, but is chosen by us for didactic reasons, we give here a short account of the way in which the fundamental equations appear in the practical system (the rationalized MKS system).

These units have three definite features. The first is that length and mass are measured in terms of meters and kilograms. This brings many basic measurements closer to human experience. The second feature is the assignment of definite constants, permittivity κ_0 and permeability μ_0, to free space. It is this which renders the units difficult from the point of view of basic instruction. The third is the rationalization of the equations to render them simple. This is done by sacrificing simplicity in the statement of Coulomb's and Ampere's laws to gain simplicity in the derived equations.

With this background it can be seen how the MKS system operates. The following equations hold.

Coulomb's law is

$$F = \frac{1}{4\pi\kappa_0}\frac{e_1e_2}{r^2}$$

F is in newtons, and

$$\kappa_0 = 8.85 \times 10^{-12}\,\frac{\text{coulombs}^2}{\text{joule meter}} = 8.885 \times 10^{-12}\,\frac{\text{farads}}{\text{meter}}$$

κ_0 is analogous to the dielectric constant of free space and is called the *permittivity*.

Energy per cubic meter $= \dfrac{\kappa_0}{2}E^2 + \dfrac{\mu_0 H^2}{2}$ (E in volts per meter and H in amperes per meter)

Ampere's law is

$$dB = \frac{\mu_0 i\,dl \sin\theta}{4\pi r^2}$$

μ_0, the permeability of free space, is 1.257×10^{-6} henries per meter; i is in amperes. Since the field in free space contains μ_0, it is written as **B**; **H** is really a constructed quantity. No moving electron ever feels **H**. This is a definite point of difference between the MKS system and the Gaussian system.

The force on a current element is

$$\mathbf{dF} = i\,\mathbf{dl} \times \mathbf{B}$$

Maxwell's equations are

$$\text{curl}\,\mathbf{H} = \mathbf{i} + \frac{\partial\mathbf{D}}{\partial t}$$

$$\text{curl}\,\mathbf{E} = -\frac{\partial\mathbf{B}}{\partial t}$$

The energy flow across a square meter per second is $\mathbf{E} \times \mathbf{H}$ watts.

It can clearly be seen that the final equations are neater. There has been some sacrifice of simplicity in the statements of the basic laws, and the meaning of κ_0 and μ_0 is very hard to visualize. For general use, practical units are to be commended as they more readily offer comparison with measurements. The reader should have no difficulty in operating in any system he chooses. A comparison between units is given by Spees,[1] which may prove helpful.

[1] A. H. Spees, *Am. J. Phys.* (1947).

AUTHOR INDEX

SUBJECT INDEX